AUTOMATED DEVELOPMENT OF PROCESS TIME ESTIMATING MODELS

Taqui Hassan Ansari Shaik

A thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of the De Montfort University

November 2006

De Montfort University

Abstract

This research has examined the cost estimating and cost modeling research literature and identified the benefits and limitations of existing practices. Particular emphasis has been placed on the methods available for developing cost models at the early stages of product and process development where data from which to develop models is scarce. Shortfalls in existing practices have been identified as well as potential methods of resolving these limitations. Of these methods Virtual Manufacturing appears to offer the greatest potential for resolving issues with lack of data availability by enabling such data to be generated.

Detailed trials have, therefore, been undertaken to examine the effectiveness of Virtual Manufacturing in terms of its ability to generate valid data in the quantities required to ensure accurate cost models can be developed. In addition, the research has involved the use of Data Mining techniques to identify the cost estimating relationship's within the data output from the Virtual Manufacturing trials. Here the aim has been to investigate the potential use of Data Mining techniques to fully automate the data analysis stage of the cost model development process.

Acknowledgements

I would like to express my thanks to Professor David Stockton for his support and guidance during the course of this study at De Montfort University.

I would also like to thank my second supervisor Dr. Zulf Khan and my colleagues Dr. Khalil, Dr. Steeple and Mr. Khalifa for their support and motivation. They have each contributed valuable suggestions and advice that has assisted me in completing the research.

I wish to acknowledge the PhD studentship received from the De Montfort University which enabled me to undertake the research. I am also grateful to the industrial collaborators involved in this project.

I would like to thank my Parents, wife, brother and my friends for their love and support. Also I would like to thank both of my wife's and my families who believed in me and supported us through out this exceptional time.

Declaration

I declare that the work described within this thesis was originally undertaken by myself, (Taqui Shaik) between the dates of registration for the degree of doctor of philosophy at De Montfort University, October 2002 to November 2006.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 5

Chapter 6

Glossary

ae	depth of cut
ANNs	Artificial Neural Network
B_s	Batch size
BS	British Standard
C_{tm}	Painting Cycle time
CAD	Computer Aided Design
CMD	Cost Model Development
CNC	Computer Numerically Controlled
CERs	Cost Estimating Relationships
CS	Continuous Simulation
D	Milling tool diameter
dt	diameter of the cutter
d_1	diameter of the cutting section
d_2	diameter at opposite end being cut
D_c	Depth of cut
Dwp	Distance between paint gun and work piece
DBCC	Drilling Burr Control Chart
DES	Discrete Event Simulation
DOE	Design of Experiments
F_t	Feed per tooth
FD	Find Dependencies
FL	Find Laws
FMAUT	Fuzzy Multi-Attribute Utility Theory
G_s	Paint gun speed
G_{rg}	Paint gun range
GMDH	Group Method of Data Handling
IDEF	Integrated Definition

IPV	Independent Process Variables
l_1	full length of the tool
l_2	tool cutting length or flute portion
L_c	Machining length
lw	length of work piece
MAPE	Mean Absolute Percentage Error
MIV	Most Influencing Variables
MIP	Model Identification Process
MOST	Maynard Operating System Time
MPM	Mathematical Process Models
MTM	Method Time Measurement
n	Spindle speed
n_o	Number of operations
n_t	Number of tools
OLAP	On Line Analytical Processing
OA	Orthogonal Array
P_{fr}	Paint fluid flow rate
P_{tl}	Path length,
P_{ns}	Paint sprayed, and
P_{nw}	Paint wasted
PnP	PolyNet Predictor
PMTS	Predetermined Motion Time Systems
r_e	Proportion of material removed by external machining
r_i	Proportion of material removed by internal machining
RTM	Robot Time Motion
SLR	Stepwise Linear Regression
SKAT	Symbolic Knowledge Acquisition Technology
T	Turning Cycle time
T_m	Milling Process time
t_{ln}	Loading and unloading time

t_{pt}	Tool positioning time per operation
tm	milling machining time
t_{sa}	Basic set-up time for machine
t_{sb}	Set-up time per tool
TE	Transfer Efficiency
V_c	Surface speed of tool
V_f	Feed speed
VM	Virtual Manufacturing
VDG	Video Data Generation
z	number of flutes in a tool

Chapter 1 Introduction

1.1 Manufacturing Cost Modelling

In order to compete within increasingly competitive markets (Layer, 2002; Salas, 2004; Wooding, 1997; De Rosa, 1999), the manufacturing organisation needs to provide;

i. greater choice of products

ii. greater amount of product customisation

iii. greater choice of materials

iv. greater choice of manufacturing processes

v. reduce product development cycles

vi. high return on investment products

There often exist hierarchical relationships amongst the above listed factors. For example greater choice of products can results from the availability of greater choice of materials, and similarly greater choices of manufacturing process may result from greater choice of materials. Wang (2000) identifies that in order to "support the product and process development needed to meet these market requirements, it is expected that quantity, type, accuracy and complexity of cost information to be generated will need to be greatly increased." Cost models will play an increasing greater role in helping to provide this cost information. In this respect greater numbers of cost models will be required to be able to represent greater levels of product and process complexity. The emphasis on innovative products and processes will mean that there is less historical cost data available (Salas 2004) from which to generate cost models and less product and process expertise to guide the cost model development process. Lack of historical data, complexity of product and processes and lack of product and process expertise will make it difficult to undertake, effectively, the main tasks involved in developing cost models, i.e. data identification, data collection and data analysis.

Of particular concerns is the data identification and data collection tasks which through lack of historical data and process expertise will significantly effect the time, cost and accuracy of the cost model development process. In addition, although there are a range of advanced data analysis techniques available, their effectiveness within the cost model development domain remains largely unknown. With the insignificant levels of product development necessary to compete the ability to estimate accurately cost at the concept stage of development process will become essential. Over the past 30 years the prevalent cost estimating techniques used for the development of cost models at the conceptual development stage, discussed by Fisher (1974), have been the basic bottom-up, analogy and parametric approaches. Most of these methods rely heavily on expertise and historical data. This can be a major limitation as historical data or the relevant expertise may not be available in either their entirety or in a reasonable time frame.

1.2 Research Aims and Objectives

1.2.1 Aims of the Research

The aims of the research are to overcome the above limitations of the existing cost modelling approaches by generating improved methods for developing cost models for products and processes that are in the conceptual design stage of their development and commercialisation.

1.2.2 Research Objectives

i) To establish a suitable method to generate data for the purpose of cost model development process.

ii) To determine the most appropriate data analysis techniques for use within the cost model development process.

iii) To evaluate the effect of complexity in the manufacturing process in terms of numbers of variables and numbers of data points.

iv) To reduce the time taken to develop cost models through automation of the modelling process.

1.2.3 Research Methodology

i) To investigate the use of data generation techniques e.g. Discrete Event Simulation, Continuous Simulation, Virtual Manufacturing, Mathematical Process Models and Video Data Generation.

ii) To validate the use of selected data generation technique.

iii) To identify suitable data analysis techniques to support the development of cost models.

iv) To investigate data analysis techniques to facilitate the cost model development process e.g. Linear Regression, Artificial Neural Networks, Fuzzy Logic and Data Mining.

v) To identify data analysis techniques to assist in building accurate CERs especially with large number of data points and variables.

vi) To identify the different level of process details and the number of process variables available from the process models.

vii) To identify the time required to build virtual process models.

viii) To develop a novel approach to assist in the process of CMD using the above two identified modelling techniques.

1.2.4 Scope of the Research

In order to improve the cost model development process for manufacturing processes which are at their conceptual design stage, the following summarise the scope of this research:

i) research current knowledge and best practice regarding the cost model development process,

ii) research new methods of data generation,

iii) research new methods of generating cost estimating relationships,

iv) Undertake a programme of experiments to derive these new methods,

v) select the most appropriate method,

vi) validate the selected method against a simple mature process where modeling and results are well correlated e.g. milling,

vii) apply the selected approach to a complex manufacturing process such as Automated Paint Spray,

viii) evaluate complexity issues and validity of cost models generated,

ix) automate the process of cost model development.

1.3 Structure of Thesis

Chapter 2 starts with an overview of the basic tasks involved in the cost model development process and use these to discuss the characteristics of the traditional cost model development process. Individual cost model development tasks are examined and their limitations are identified. Essentially these tasks are found to be primarily data identification, data collection and data analysis activities. Analysis of the cost modelling research reveals that in terms of cost modelling at the concept design stage the shortcomings of the existing methods are in the areas of data identification and data collection. This is found to be due to lack of historical data sources from which appropriate data can be collected. In addition, there is a general problem with existing data analysis methods which is attributed to limitations in their ability to deal with large numbers of predictor variables and the types of relationships between these data types.

The purpose of Chapter 3 is to examine alternate data generation methods that could assist in overcoming the limitations of existing methods. This chapter therefore begins with identifying, and where necessary modifying, the criteria for the selection of a suitable data generation method at the concept design level. These criteria are then used to compare alternative data generation methods, which include continuous simulation modelling, discrete event simulation modelling, virtual manufacturing, mathematical process modelling, pre-determined time standards, and video data generation. From this analysis virtual manufacturing is found to be the most suitable potential candidate.

This chapter also examines alternative data analysis tools currently used for the purpose of cost model development. Here regression analysis is compared with advanced data analysis techniques, which include artificial neural networks and fuzzy logic. The review highlights the shortcomings of these techniques within the cost model development process. Particular shortcoming of existing processes is their reliance of manual input to select predictor variables, which prevents the further automation of the data analysis process. This section also examines a further advanced data analysis tool, i.e., data mining which when compared to the other examined appears to posses greater potential for enabling the cost model data analysis process to be automated.

With selection of the Virtual Manufacturing and Data Mining techniques in the previous chapter, Chapter 4 now provides details of the experimental plans for testing the feasibility of use of these techniques within the cost model development process. Areas of manufacturing for their application have been selected for testing the Virtual Manufacturing technique, i.e. Vertical End Milling and Automated Paint Spray. Virtual manufacturing process models of these two processes have then been built capable of performing machining and spray-painting operations respectively.

In order to test the Data Mining technique, Taguchi's Orthogonal Arrays have been used to design suitable experimentation. These experiments were designed to analyse the data generated from these virtual process models using data mining algorithms that include Stepwise Linear Regression, Find Laws, PolyNet Predictor and Find Dependencies. Further experiments were also undertaken to test the effect of "number of variables" and "number of data points" on the estimating accuracy of the data mining process using an additional data set for the turning process.

Chapter 5 presents the cost models developed using the data mining algorithms and details of estimating their accuracies. Additional methods have also been tested for their ability to improve the estimating accuracy of the models output from the data mining process.

Chapter 6 discusses the key concepts of the research and summaries the overall approach and research methodology. Chapter 7 provides the research conclusions and Chapter 8 highlights further work.

Chapter 2 Cost Modelling

2.1 Introduction

Cost models have been developed using a wide range of approaches of which the primary ones are parametric models, top-down models, bottom-up models and subjective judgement. This chapter describes the cost model development process in terms of:

a) the characteristics of the individual tasks involved, and

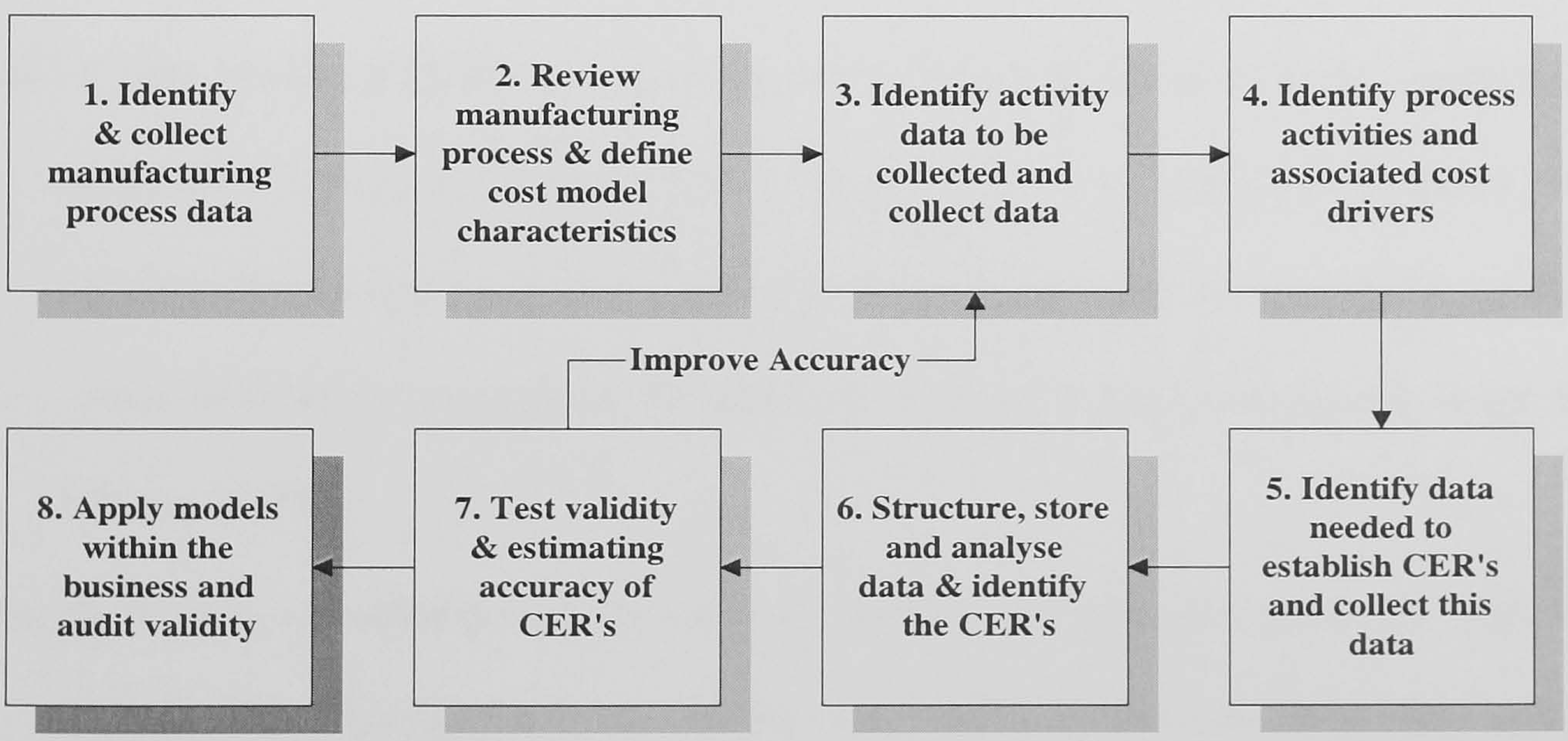
b) the characteristics of the cost models to be developed.

2.2 The Cost Model Development (CMD) Process

2.2.1 Characteristics of the CMD process

Traditional CMD methods, which are still primarily used today, are based on the sequence of processes shown in Figure 2.1.

Figure 2. 1 The Cost Model Development Process (Stockton and Wang 1999)



Within the basic CMD process shown in Figure 2.1 there are three main types of task, i.e. data identification, data collection and data analysis (Stockton 1999). The characteristics of the traditional CMD process are:

i. It is an iterative process as indicated by the 'Improve Accuracy' feed back arrow in Figure 2.1. This iterative nature ensures that an acceptable estimating accuracy is aimed for and leads to repeating questions such as:

    a. have the right product and process variables effecting costs been identified,

    b. within these product and process variables, have the right cost drivers been selected, and

    c. have the best cost relationships been identified.

ii. It normally requires high levels of process and /or product expertise, i.e. the selection of the right type of process and product data requires subjective judgement by personnel with high levels of product and process expertise. Unavailability of such levels of expertise may lead to inaccurate or invalid cost models.

iii. It may involve collecting data at varying levels of detail. Hence, the availability of valid data sources is a further factor, which can have a significant influence on the quality of the resulting model. Lack of reliable data sources is a particular constraint when developing cost models for processes that are at their conceptual design stage (Wang 2000).

iv. It is time consuming, i.e. in order to achieve the required accuracy with valid relationships amongst the cost drivers, the CMD process may need to be repeated many times.

v. Cost models may be difficult to validate due to a lack of available data that relates cost to cost drivers.

vi. In general, the personnel involved in the traditional CMD approach are restricted to process experts with little structured inputs from other functions.

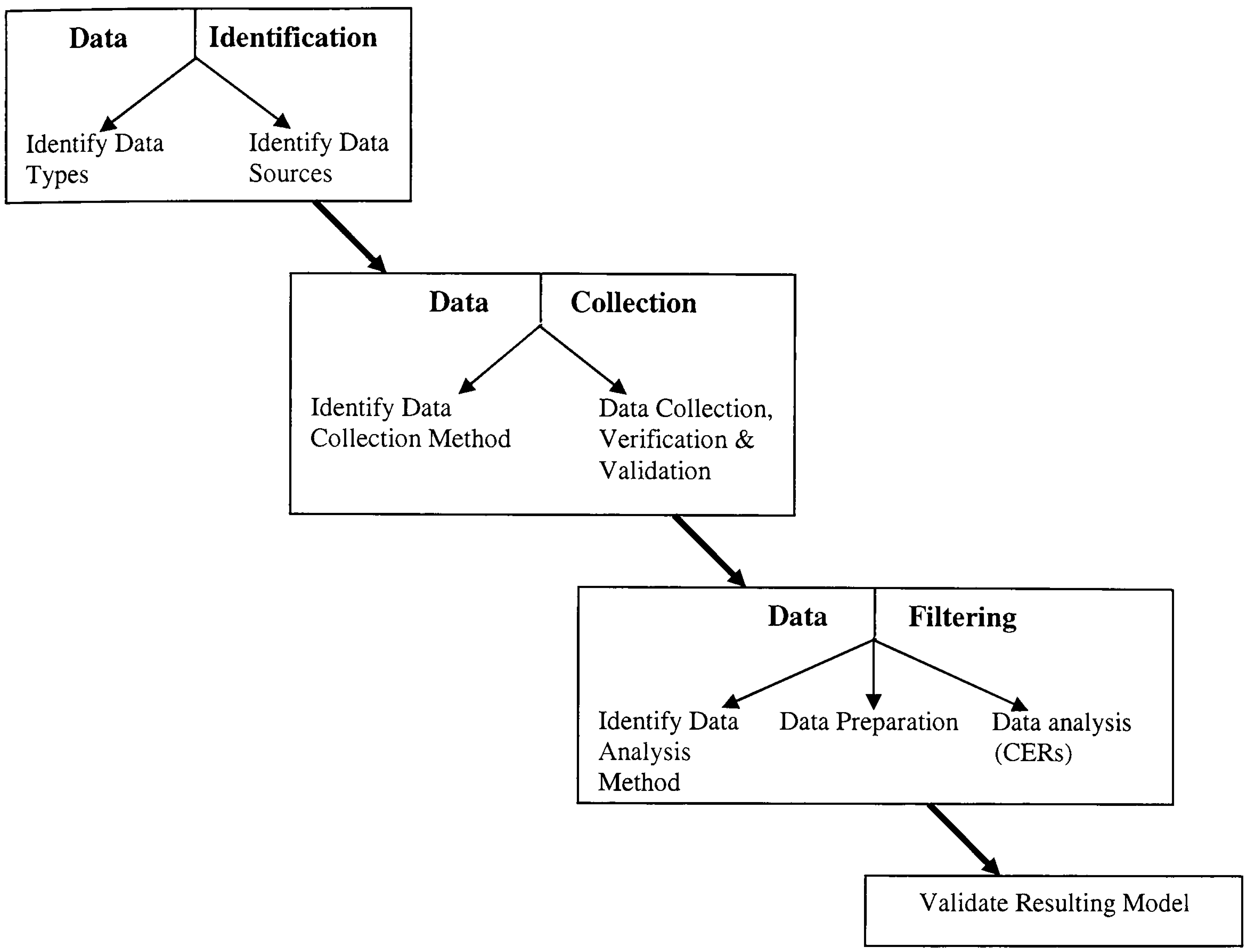
vii. The correct identification of valid cost drivers is dependent on the level of experience of the personnel developing the cost models and the absence of bias in their selection of cost drivers.

viii. There is often a dilemma between the need for specific cost model characteristics, e.g. level of accuracy, and the availability of data to ensure these characteristics can be achieved.

2.3 Cost Model Development (CMD) Tasks

Figure 2.2 shows the hierarchy of CMD tasks in terms of these basic activities with each basic activity divided into their respective sub-tasks. Essentially the cost model development process starts with identifying potential cost drivers and then continues with the collection of relevant data for these cost drivers. The final stage of the process is to analyse the collected data in order to identify the relationships between costs and cost drivers.

Figure 2. 2: CMD Tasks



2.3.1 Identifying the Types of Data that need collecting

The primary aim of this task, (Wang, 2000; Agpar, 2003), is to identify the potential independent predictor variables of the cost estimating relationships. The basis of developing accurate estimating models starts with identifying the right data types (Randal 1996), i.e. the cost drivers. According to Delgado (2002) four basic data types need collecting, i.e.:

a. product features,

b. process features,

c. process activities, and

d. cost resources types, which are made up of labour, material, and process times.

A structured approach to the traditional CMD process has been developed by Delgado, McNeill and Stockton (2002) i.e., COSTMOD methodology. This methodology attempts to structure the task of data identification and data collection using two basic stages, i.e. a Process Scoping stage and a Model Identification Process (MIP). The objective of the Process Scoping stage is to identify the characteristics of the model being developed and the levels of product and process data that need to be collected. The MIP phase involves an initial data collection task and the use of this data to identify the specific information required to develop the cost model. The MIP stage consists of four basic tasks, i.e.

i) identifying and collecting relevant data, i.e. process features, product features and process activities,

ii) prioritising data on the basis of their effect on process cost,

iii) comparing the prioritised data in order to identify potential relationships that exist amongst them, and

iv) identifying the most significant of these relationships for prioritising data analysis tasks.

There are a wide variety of data types that could be considered for inclusion within a cost model. For example, Dauda (2005) identified machine, tools, labour, energy and coolant as major cost drivers for a high efficiency deep grinding process. In sheet metal operations the product features, used within cost models included sheet thickness, number of corners, number holes and inner and outer length (Verlinden 2005). Curran (2003) has considered a geometrical complexity factor, manufacturing complexity factor and specification complexity factor as major data types to be used in conceptual cost modeling for aircraft manufacturing. According to work undertaken by Rush and Roy (2001) cost drivers product characteristics can include:

a. geometry of the product to be manufactured,
b. material of the product,
c. production process selected to manufacture the product, i.e. since this determines the resources required for production, and
d. production planning.

Inclusion of all factors, that may contribute to process or production costs, was found by Needy (1998) to hinder and prolong the data collection task. Therefore, the personnel involved in the CMD process should ensure that the data types identified closely match the required cost model characteristics in terms of accuracy, time to develop cost models and the level of detail of the estimates expected from the cost model.

According to Rush and Roy (2001a) process maps can be used as a means of identifying cost drivers. This research used the IDEF methodology to develop the required process maps using an integrated product team. According to Weustink (2000) product cost at the early design stage can be estimated but cannot be controlled. This is due to the unavailability of detail product cost driver information at these conceptual stages of design. This leads to less accurate and, therefore, less reliable cost estimates particularly for standard product ranges (Perera 2001).

Jung (2002) presented a feature based cost estimation system for machined parts. The proposed system addressed the issue of early cost estimation using manufacturing process, (i.e. machining) and product features. The machining features were categorized based on the product geometric features such as rotational or prismatic. Features such as "machining areas" were extracted from product drawings and respective machining time was calculated using mathematical formulae. Finally, other variables such as, setup time, material cost and overheads, are obtained from historical data.

Rush and Roy (2001b) examined the use of expert judgment in manufacturing cost estimation. According to Roy cost estimating is a process, which greatly relies on the expert knowledge of the estimator. To overcome this subjective nature, the research focused on capturing the rational and knowledge underlying the cost estimating process. Roy concluded that the use of computer based cost estimation may reduce the level of reliance on expert judgment but cannot replace the process as a whole and where reliance on subjective expertise is necessary the accuracy of estimates cannot be guaranteed.

2.3.2 Identifying Relevant Data Sources

Wang (2000) provides a comprehensive list of the potential data sources applicable to the development of cost and process time estimating models i.e. Table 2.1. The sources from which data is derived effects both the accuracy and the validity of the resulting cost model (Hamilton 2002). Development of cost models requires reliable and timely data sources to be found. If relevant data sources are not available then cost modelling is ineffective.

Table 2. 1: List of Data Sources for CMD

Actual Process	Operator's Black Book
Video of Process	Process Controllers
Costed Components	C A D Files
Process Expert	C N C Program
Training Manuals	Equipment performance
Operating Manuals	Empirical Laws
Process Models	Synthetic Standards
Maintenance Manuals	Departmental records
Shop floor	Quality Manuals
Documentation	Planning & Control
WWW	Sheets
Patents	Literature reviews
Similar Process	Physical Models
Equipment Specification	

The characteristics of data sources of importance to the development of valid CER's include:

a) availability of data,

b) level of detail of data,

c) accuracy of data,

d) amount of data, and

e) validity of data.

Easy availability of data sources helps in minimizing the time required to build the cost models. Although, there are many locations where data can be collected from it is important to identify reliable sources. The amount of data available to build a cost model depends on the stage at which the cost model is being developed, i.e. conceptual stage or detail designed stage.

Roy (2001) carried out a study to develop CER's for aiding the design process using a multifunctional team in an integrated product development environment. This methodology enabled the inclusion of both quantitative and qualitative predictor variables. The CERs developed were validated with the help of case studies. The research identified the need for improved data collection methods for qualitative types of data.

Wong (1992) proposed a computer integrated manufacturing cost estimation system, which incorporates automated generation of cost estimates using design input. The model is expected to extract information from different data sources such as CAD, process plans, production plans, bills of material and production time studies. Data collected from these sources provided material costs, labor costs, production costs and overhead costs which were then summed to generate the overall product cost. Mileham (1993) used three types of data sources, including a CAD system, in the development of an estimating system that made use of parametric cost models. Here data sources included conceptual design data, process information and process variables, which are capable of describing component characteristics.

2.3.3 Identify Data Collection Methods

Once the data types and their respective sources have been identified, it is necessary to decide the method by which the data will be captured. The various data collection methods that can be used to aid this process have been identified (Currie, 1977; Stockton, 2000; Scanlan, 2002; Wang 2000), and are:

a) work measurement techniques such as MTM and MOST,

b) brainstorming and group discussion,

c) sampling,

d) flow charts,

e) questionnaire and surveys,

f) direct observation of processes,

g) discrete event and continuous simulation,

h) virtual manufacturing environments,

i) process modelling, and

j) video recordings.

2.3.4 Data Collection, Verification and Validation

Once identified the data collection method must be used to actually collect data and ideally this data must contain all the variable types that influence the cost being estimated. According to Baguley (2004) data collection is a time consuming process and is constrained by the following factors:

a) availability of data source,

b) amount of data to be collected,

c) level of detail of the data, e.g. detailed process data or abstract data,

d) consistency in the data collection process, and

e) frequency of data collection.

A verification check must be undertaken to ensure that data is correctly transferred to the required input data storage medium that forms part of the cost modeling process. In the verification stage, the data is checked for manual errors that may have occurred whilst collecting the data. Another key application of this stage is to identify and to assure that sufficient data has been collected to obtain the correct cost model characteristics. A validation check therefore, is made to ensure that all data entered is sensible.

2.3.5 Data Preparation

The data preparation stage involves removing outliers from the collected data, and identifying and categorising quantitative and descriptive values. This task should be carried out bearing in mind that the accuracy of the model obtained from the data analysis tools is heavily reliant on the data set used in the analysis. The "cleaned" data sets should make data analysis easier to perform. Later in the data preparation process this data can then be transformed to a form acceptable by the data analysis tool.

In detail the data preparation process includes the following steps:

i. Transforming the data to a form acceptable by the chosen data analysis tool. Table 2.2 for example lists the different data types which typical data analysis algorithms can accept:

Table 2. 2: Data Analysis Algorithms and Data Types

Algorithms	Data Types
Stepwise Linear Regression	Numerical
Find laws	Numerical, Categorical
Artificial Neural Networks	Numerical, Categorical, String
Classify	Yes / No
Decision Tree	Numerical, Categorical or Boolean (Yes/No) attribute
Nearest Neighbor	Numerical, Integer, Category, String, Yes/No

ii. Transforming the data to facilitate increased data analysis speeds. This involves interpreting the data in order to extract the information from the data set. For example, if a data set includes "year of car production", (e.g. 1970), it normally refers to "date" and not the number "1970". This data could be interpreted and/or transformed to represent the age of the car where its year of production is compared with the present year, eg. 2006-1970 is 36 years of age. The process of interpreting data sets in this manner is often necessary at the early stages of data analysis.

iii. Removing data columns that may not contribute to the analysis. For example, if all values of a variable are equal it would have no correlation with the dependent variable. Hence the column containing this data is of no value to the analysis process and would, therefore, be removed.

iv. Merging separate collections of data into one dataset. For example, shop floor information could exist in a number of places, e.g. display boards and log sheets. It would also be held in different data formats. This unorganized information cannot be used directly for analysis and therefore needs merging.

v. If the amount of data available is too great for the data analysis software to handle, it may be split into sections so as to better fit software size limitations.

vi. Extracting samples of collected data may be necessary in order to analyse under different scenarios. For example, a data set is often split into two parts, one part being used to perform the analysis and develop the cost models and the second part being used to verify the accuracy of the models generated.

2.3.6 Select Data Analysis Tool

The aim of this task is to identify the best data analysis technique for the purpose of establishing the CERs. There are a variety of data analysis tools available which have been used in the process of cost model development. The need for selection of the data analysis tool is to ensure that the selected tool can deal with the amount of data points available, the type of relationships that exist and number of variables required within the model. MacKenzie (2003) identified that with increasing numbers of data points and predictor variables the effectiveness of traditional data analysis techniques such as regression analysis quickly deteriorated, particularly the levels of estimating accuracy that could be expected.

2.3.7 Analyse Data to Develop Cost Estimating Relationships (CERs)

Having identified the cost drivers, the relationships they have with the cost of the process, (i.e. the dependent variable), have to be established. Again the extent of relationship between the process or product cost and process independent variables depends on the objectives of the cost model. For example, if the impact of part design on production costs is to be investigated then the CER should reflect the cost effects of such variables as part size and design features. Similarly, to identify the effect of changes in volume and production mix then the CER needs to contain variables that account for these effects. In most of the manufacturing organisations the cost is directly related to the cycle time of the process, which is the time taken to complete the whole process. Therefore, in the CER development process, cost equations are often developed based on estimating process times.

CERs range from simple rules of thumb to complex non-linear mathematical relationships involving multiple variables. To cope with this variety, a range of data analysis techniques have been applied in the development of CERs including a variety of statistical and advanced modelling techniques which include:

a) mathematical models developed using empirical data,
b) various forms of regression analysis, i.e., linear regression, multiple linear regression, stepwise linear regression and non-linear regression,
c) fuzzy logic, and
d) parametric modelling.

Examples of the application of these methods are briefly described below:

In terms of mathematical models, Kim and Dronfeld (2001) developed a cost estimation procedure for a drilling operation. The cost of drilling was divided into two parts, i.e. cost of hole making and cost of deburring. Initially, a Drilling Burr Control Chart (DBCC) was developed using Bayesian statistics to represent the distribution of burr types in the form of mathematical formulae. These formulae define the distribution of burr types in terms of process variables which themselves were related to process costs.

Stockton (1982) presented a series of cost estimating models developed using regression analysis for application within low volume / high variety machine shops. The research demonstrated versatility of regression analysis in its ability to produce models for a wide range of manufacturing processes. The work also demonstrated that using multiple regression analysis quick estimates were possible when compared with traditional detailed estimating methods. Schreve (1999) also demonstrated the significant role that manufacturing cost estimation models could make in the design for manufacture process. Here cost models for fabricated steel assemblies were developed using time studies to collect data and regression analysis to develop CERs from this data. Inter-company comparisons were used to validate the use of the models for design for manufacturing purposes.

Popham (1996) described parametric modelling as a methodology suitable for processes where reliable historical records of similar processes were available. It was found that the quantity and quality of data is of most importance in these types of estimates. According to Roy (2000) parametric costing can be used at all stages of the product development process. However, it is primarily used in the early stages of product or process design where sufficient quantitative data is available.

Wang (2001) examined the application of Artificial Neural Networks (ANNs) to the CMD process in terms of their ability to identify valid relationships amongst cost drivers and the levels of estimating accuracy of the resulting ANN models. This research developed a method of determining which combination of ANN structural elements provided the highest level of estimating accuracy. Using this methods ANN based models were developed that outperformed multiple linear regression based models in terms of estimating accuracy. In addition to the ANN architecture, it was found that both the number of process variables and the number of data points have significant impact on the estimating accuracy of the resulting models.

Fuzzy logic based cost models have been developed by Shehab and Abdullah (2002) and Baguley (2004). Shehab and Abdullah presented a fuzzy logic rule based system for product cost modelling for use at the early stages of product design. Fuzzy logic was used to estimate costs and to deal with any uncertainty associated with the cost model. The CMD process developed by Baguley (2004) follows the data identification and data collections concept developed previously by other researchers (Delgado, 2002, Stockton, 2001, and Wang, 2000). However, this research considered the use of fuzzy logic as a data analysis tool. The research adopted a similar approach to that of Wang (2001) in examining the architecture of data analysis tool. It compared cost models using alternative fuzzy logic structures with multiple linear regression based models and found that the use of fuzzy logic was capable of estimating costs to an appropriate accuracy level suitable for commercial use i.e. accuracies of 6%. The research proposed a decision-making methodology based on the use of Taguchi Orthogonal Arrays to select the best fuzzy logic architectural elements. The research also extended the range of fuzzy logic methods used within the CMD process to include the Takagi Sugeno Kang method, Adaptive-Neuro-Fuzzy-Inference and subtractive clustering method.

2.3.8 Determining Model Accuracy

Once the CERs for a particular manufacturing process have been developed, it is then necessary to validate these models. In addition, the product and process assumptions under which models have been developed must be formalised. A traditional method of validating CERs is comparing the resulting estimates with the actual costs, i.e. determining if they have acceptable estimating accuracy. Mason (1997) points out that it is the combined use of the model with the resulting overall estimating process that must be validated not merely the application of the model itself.

Validation of a cost model implies that the accuracy of the model is within an acceptable range and can be used as an estimating tool. However, there would still remain a degree of uncertainty attached to each predictive model that is being developed. This could be due to a lack of a suitable data source and the factors discussed in Section 2.3.2. Moreover, there are no fixed boundaries to define the estimating accuracy acceptance level that can be achieved by a model prior to its development and according to Scanlan (2002) there is no predefined methodology for defining the accuracy or associated uncertainty with the resulting cost estimates. Therefore, a degree of uncertainty and a measure of statistical significance are essential parts of the model building process, i.e. risk analysis is now becoming an essential step in the cost model development process. Validation of the model, is not formally undertaken, in terms of ensuring that:

i. there is an actual causal relationship between each predictor variable and the dependent variable,

ii. the value of each variables coefficient is valid in terms of both its sign (+ve or –ve) and is relative size, i.e., accurately represents variables actual affect.

2.4 Summary

Accurate and consistent cost estimates are always a distinct advantage in this rapidly changing global market. To meet this objective, changes must be made to conventional cost estimating practice. Research concerning traditional costing system revealed that, they required model builders and users to have knowledge and experience of the different stages involved in the development of cost models as well as processes being costed. In addition, according to Scanlan (2002) current cost modeling tools lack key features such as adequate visualisation and presentation of the resulting cost data and integration with advanced CAD tools.

Wang (2000), Roy (2001a), Stockton (2001), and Delgado (2002) describe the tasks such as data collection, data identification and data analysis are both labour intensive and time consuming. In addition, Whiteside (2003) adds that speed, accuracy of the CMD process, levels of risk and confidence involved in using estimates are the areas of the cost model development process that need attention.

Further investigation of these problem areas of cost model research lead to the following inferences:

i. at the early design stage there is generally lack of sufficient data from which to develop cost models with the required accuracy levels. Hence, there is a need to generate such data.

ii. problem of existing data analysis such as regression analysis is attributed more towards it's failing to deal with more number of variables and data points to develop more accurate CERs.

iii. existing data analysis techniques such as regression analysis and neural networks are found to be restricted in their use, i.e. in terms of their ability to deal with different type of relationships between dependent and independent variables and output from these techniques. Therefore, there is a need to use advanced data analysis tools to overcome the shortcomings of these existing techniques.

Chapter 3 Data Generation and Analysis

3.1 Introduction

Chapter 2 examined the main practices involved in the cost model development process. Several areas for improvement were identified in these practices, i.e. a lack of process data to develop cost models and a deficiency of existing data analysis techniques to handle large amounts of data and/or predictor variables. The purpose of this chapter is, therefore, to examine in detail these two aspects with the aim of identifying methods of resolving current limitations.

Where insufficient historical data exists, the practice of "generating" data will be examined. This chapter, therefore, begins by examining alternative methods of data generation which include the use of:

a) continuous simulation modelling,

b) discrete event simulation modelling,

c) virtual engineering,

d) mathematical process modelling,

e) pre-determined time standards, and

f) video data generation.

This chapter will then compare the use of a traditional data analysis technique, i.e. regression analysis and advanced data analysis techniques such as artificial neural networks and fuzzy logic with alternative data analysis process i.e. data mining.

3.2 Comparison of Data Generation Methods

In order to develop a set of criteria by which alternative data generation methods can be compared the CMD process characteristics identified as part of the COSTMOD methodology (Delgado, Stockton 2002) were initially examined. These characteristics were designed for a cost model development process that relied primarily on the collection of historical data or the use of process experts where no such process history existed.

The process scoping characteristics identified in the COSTMOD methodology (see Table 3.1) were, therefore, modified to ensure their relevance to situations where data needs to be generated.

Table 3. 1 Process Scoping Characteristics of COSTMOD Methodology

<table>
<tr><td rowspan="4">Model Validated and Verified by</td><td>Model Owner</td></tr>
<tr><td>Model Developer</td></tr>
<tr><td>Model User</td></tr>
<tr><td>Model Function</td></tr>
<tr><td colspan="2">Time available to collect data</td></tr>
<tr><td colspan="2">Time available to input data</td></tr>
<tr><td colspan="2">Resource/Hours Available to collect data</td></tr>
<tr><td colspan="2">Resource/Hours Available to input data</td></tr>
<tr><td colspan="2">Resource/Hours Available to develop the CER</td></tr>
</table>

This modification produced the following set of characteristics, i.e.:

a. time required to develop initial data generation model,

b. time required to carry out data generation experimentation, and

c. time required to collect the generated data.

These modified characteristics will be used within this chapter to compare the effectiveness of alternative data generation methods. When comparing methods it is also necessary to establish for each alternative method the types of data that can be generated. Here, Delgado, Stockton et al (2002) identified the basic types of data within cost models as:

a. direct and indirect resources that need to be estimated, i.e. material, equipment and/ or labour costs,

b. direct and indirect process times that need to be estimated,

c. product features of which there are three levels i.e. product level (engine), assembly level (piston and connecting rod) and component level (connecting rod),

d. process activities of which there are three levels i.e. processes (milling), operations (set up and machining) and operational activities (machining).

e. process features of which there are three levels i.e. process level, process sub-assembly level and machine level. Although Delgado et al (2002) makes use of three levels, the current research views process features not as 'predictor variables' but as 'constraints'. For example, the value of the process feature 'feed rate' needs to be set such that the correct level of machining quality may be derived and therefore acts as a process constraints.

It is also essential that the data generation method selected produce data that links the resources or times being estimated by the model, (i.e. a and b above), with one or more of the predictor variable categories, i.e. product features, process features and/or process activities.

3.3 Data Generation Methods

The following data generation methods have been identified and examined.

3.3.1 Discrete Event Simulation (DES)

According to Banks (1999) a discrete event simulation model can be defined as "one in which the state variables change only at those discrete points in time at which events occur." These events occur as a result of process activities starting and/or ending. Hence, "discrete event simulation involves the modelling of a system as it advances over time by representing changes as discrete events" (Heilala 1999).

DES has been used in a wide range of industries, including automotive and aerospace, and for a variety of applications. For example, Vaidiyanathan (1998) and Holst (2001) used DES to analyse existing manufacturing systems for improvements, for designing and validating new process layouts and for validating proposed capital investments. According to Law (1991) due to the ability of DES to capture time dependent behaviour it provides knowledge of activity relationships within a manufacturing system and at varying level of details (Jayaraman 1997). DES can, therefore, be used within organisations in a variety of problem solving and investment decision making processes (Banks 2000).

Levine (2004) provides details, (Table 3.2), of the types of applications DES is used for.

Table 3. 2 Application of DES

	Application Domains		
Objective \ Product phase	Process Design	Facility Design	Existing Facility
Cost analysis	√	√	√
Evaluating new technologies	√	√	√
Evaluating process options	√	√	√
Capacity analysis	√	√	√
Equipment sizing		√	
Utility sizing		√	
Scheduling			√
De-bottlenecking			√

The output data (Lawrence 2003, Jayaraman 1997, Law 1998, Ulgen 2000) generated from simulation models of potential use within cost models include data items such as process cycle times, buffer sizes, labour requirements, lead times and personnel utilisation evaluations and inventory levels, finished and raw material stock levels.

De Vin (2004) identified that when using DES 'the time to build a model' cannot be accurately predicted. In addition, DES cannot be used to analyse the model at an operational activity level, as continuous process information is normally required rather than discrete data. According to Heilala (1999) other disadvantages of using DES are:

a. the need for extensive training in model building, and

b. the difficulty in interpreting output results.

A distinct advantage of DES is, however, its ability to provide data that can be used to develop cost models that take into consideration the effect of system delays, queuing periods, inventory, idle time build-up on cycle times, lead times and labour costs rates.

3.3.2 Continuous Simulation (CS)

Continuous simulation according to Law (1992) and Gorlani (1993) involves "the modeling over time of a system by a representation in which the state variables change continuously with respect to time." Continuous simulation models describe the dynamic behavior of the system by imitating the real system over a given time period.

According to Switek (1997) continuous simulation can be used to model production activities and resources. Typical applications of continuous simulation include modeling of continuous motions of robotic arms in an automated paint spraying process. In addition, Acosta (1997) examined the application of differential equations for modeling the turning process.

One of the key advantages of this method is that it provides the real time for a process activity. However, such models are normally represented as differential equations that provide relationships for the rate of change of state variables with respect to time. According to Law (1992), simple continuous differential equations can be solved analytically, where as complex equations need methods such as Runge-Kutta integration. Such models, therefore, require a high level of mathematical ability to develop. Continuous simulators are also often slow and are only useful when using relatively simple equations with small numbers of variables, which are described, at a low level of abstraction (Law 1992). According to Baines (1999) there is normally insufficient time, expertise and/or resources available to support the widespread application of continuous simulation techniques. Although, Oscarsson (2000) suggests that continuous simulation using differential algorithms is essential where variables are constantly changing over time, in practice acceptable levels of accuracy may be possible using simpler linear models.

3.3.3 Virtual Manufacturing (VM)

According to Iwata (1997) Virtual Manufacturing is defined as "a computer system, which is capable of generating information about the structure, states and behavior of a manufacturing system as can be observed in the real manufacturing environment." A similar definition is offered by Lin (1995), who stated that virtual manufacturing "is the use of computer models and simulations of manufacturing processes to aid in the design and production of manufactured products". Virtual manufacturing involves the development of a model with specific process features and use of this model to simulate a process by using the specific process activities required to produce specific product types with specific product features. It therefore, contains all the basic data types, i.e., process features, product features and process activities, from which cost models are developed.

The typical applications of virtual manufacturing include virtual prototyping, virtual machining, virtual inspection and virtual assembly. According to Lee (2001) virtual manufacturing can be used:

a. as a knowledge (data) acquisition tool for data such as production and process data,
b. to evaluate and validate process and production plans,
c. to optimise the performance of production processes and manufacturing systems, and
d. to provide detailed information, (such as product geometry and depth of cut), about products and their manufacturing processes.

Endo (1996) used virtual manufacturing software techniques to simulate cycle times, optimise the process sequences of various sheet metal parts and for minimising the total processing time prior to actual production. This research identified good agreement between process times simulated in a VM environment and the actual times occurring in the real system. In this work, VM was used as a tool to model the cost of product development and production process as well as to generate information related to the process. Worn (2000) has identified the application of VM at different process levels in a manufacturing organisation from basic manufacturing activities, material processing, plant and facilities layout to activity co-ordination and control. Table 3.3 summarises the functions of VM models in various application areas (Rohrer 2000; Jayaram 1997).

Table 3. 3 Functions of Virtual Manufacturing

Area	Functions of VM
Design	To evaluate the product, visualise and to improve the product design
Operations management	To understand the process sequence (i.e. series of operations)
	To verify and validate process models
Manufacturing processes	To identify machining times and costs of specific products
	To predict manufacturing process capability and costs

The task of building a virtual model can be carried out using advanced virtual manufacturing software such as DELMIA (Mujber 2004). According to Kibira (2002) construction of virtual models using such software is tedious and time consuming. In terms of the CMD process VM software packages do not normally provide a standard format to read and retrieve process data for cost modeling data collection purposes.

3.3.4 Mathematical Process Models (MPM)

According to Szekely (1994) mathematical process models provide a quantitative representation of physical systems, i.e. in terms of algebraic equations. According to Wood (1999) process models are the mathematical descriptions of the relevant physical laws underlying manufacturing processes. The variables in the mathematical equations represent the state of the process entities and the mathematical operators encode the interaction between these variables (Todorovski 2005). Types of mathematical models of manufacturing processes include (Matko 1992):

a. linear and non-linear models,

b. lumped and distributed parameter models,

c. stationary and time varying models, and

d. deterministic and stochastic models.

An example of a MPM is the model developed by Boothroyd and Dewhurst (2002) for calculating the machining times of vertical milling machine operations, i.e.:

$$t_{mp} = \left(l_w + \sqrt{a_e(d_t - a_e)}\right) \cdot V_f^{-1}$$

Where:

t_{mp} = milling machining time,

l_w = length of work piece,

a_e = depth of cut,

d_t = diameter of the cutter, and

v_f = feed speed of the work piece.

Process modeling has been applied within applications such as process development, component design, and process evaluation (Wood 1999). Shatla (2000) used mathematical models to analyse and predict metal cutting process variables such as stresses, cutting forces, geometrical parameters, cutting conditions, and work piece and tool material properties. Although mathematical models of manufacturing processes have been used at varying levels of product and process design, operations planning, process control and monitoring. "The formulation of appropriate consistent models is a major, time-consuming and error-prone task" (Perkins 1996).

A disadvantage of mathematical process models is the need to use subject expertise to understand the model. Most mathematical process models are static or can only be used when steady state conditions apply. Petropoulakis (1998) further added that they are often applicable to restrictive classes of problems.

3.3.5 Predetermined Motion Time Systems (PMTS)

There are a range of PMTS systems available with the majority being variations or derivatives of the Methods Time Measurement (MTM) technique. MTM is defined as "a procedure which analyses any operation or method into the basic motions required to perform and assign each motion a predetermined time standard which is determined by nature of the motion and the conditions under which it is made" (Barnes 1980). The British Standard i.e. BS 3138, definition of PMTS is "Tables of time data at defined rate of working for classified human movements and mental activities. Times for an operation or task are derived using precise conventions. Predetermined motion time data have also been developed for common combinations of basic human movements and mental activities."

The applications of PMTS include improvement of existing processes, to guide product design improvements, to estimate product or process times and to generate standard process time data (Clark 1977). Use of such systems can aid in identifying waste, estimating process time and generating standard time data, which can then be used to improve productivity and/or process designs (Kilgore 1997). According to Kapoor (1990) the process-times data generated from work measurement systems has been used to improve budgeting processes, analyse "should-costs" and estimate delivery times. According to Barrett (1997) standard data developed at Rank Xerox helped the organisation to reduce assembly cycle times. Ince (2001) has used PTMS in order to review existing product costs and to develop a method for calculating quickly and accurately new product costs.

Cohen (1998) argues that the use of PTMS systems is time consuming because they "involve tedious calculation for conversion of shop floor data into task times". According to Luxhoj (1989) these systems are suitable for short cycle and highly repetitive tasks. Hence, they are, inappropriate for use where small manufacturing volumes or long cycle time tasks existed (Nakayama 2002). However, whilst MTM-1, for example, provides a detailed analysis of the fundamental motions involved in manual tasks the MOST technique, a simplified version of MTM-1, makes use of higher level motions and can be used within lower volume, longer cycle time work environments. In addition, computerised versions of PMTS, such as Auto MOST, can automatically generate times provided that exact descriptions of processes are supplied. They can also be applied to both high and low levels of process data (Anna 1995).

3.3.6 Video Data Generation (VDG)

Video data generation involves capturing process data by simply recording the processes on videotape. The details of a process are, therefore, made available on a screen and knowing the cost model data requirements or purpose of analysis, work processes can be observed at the is actual level of detail. In terms of the cost model development process, all data types required for cost modelling purposes can be collected, i.e. process activities, activity times, product features and process features.

Video data collection can also be used where:

a. some historical data of the process is available,
b. it is hazardous to be at the workplace while a process is operating,
c. the manual collection may cause distraction to workers and delays to operations,
d. processes are at a remote or distant location and the time and cost of visiting the site to collect data is high, and
e. operations have long cycle times and it is hence impossible to collect all data in one shift.

Although, it may appear to be a relatively simple method of collecting data it is often difficult to ensure recording is comprehensive, i.e. achieving coverage of all possible parts of the products and process, all essential 'angles' videoed, range of product types and, process features, and other process information which could contribute to the cost of the product. The captured process, either on analogue or digital tape can be analysed using windows based software such as Observer Video Pro. This software facilitates the computer-aided analysis of video recording such as human behaviour or robot arm movements in order to generate time-based data suitable for quantitative analysis (Nouldus 2003). However, the shortfall of this methodology is that the distance travelled by an object cannot be accurately measured. Hence, it is not possible to accurately determine the relationship between process time and 'distance' based process activities.

3.4 Selection of Data Generation Methods for CMD

This section compares the data generation methods identified in Section 3.3, (i.e. DES, CS, VE, MPM, PMTS and VDG), in terms of their ability to meet the process scoping and data generation requirements identified in Section 3.2. These comparisons are shown in Tables 3.4, 3.5 and 3.6 respectively, which were carried out using both qualitative and quantitative research methods, which included:

i) Process models experiments, i.e.; process models were built for DES, VM and VDG,
ii) Discussions with software vendors, i.e., Delmia and Nouldus for DES, VM and VDG,
iii) Discussions and brainstorming sessions with process experts for CS, MPM and PMTS, and
iv) Literature review for all methods.

Based on above listed research methods the following observations can be made from examination of Table 3.4, i.e. all methods can be considered approximately equal in the total time required to develop the model, undertake data generation experiments and collect the data resulting from these experiments. This can be best understood using an example, from Table 3.4, where the time required to develop the initial data generation model is one month using VM and using VDG is one week where as the time required to carry out data generation experimentation is vice versa. In this way the overall time taken to carry out these data generation tasks by these methods is approximately equal. In addition, the limitation to the times identified in Tables 3.4, 3.5 and 3.6 is that they may vary based on conditions such as model developer's expertise, complexity of models considered and the research methods used to identify development times.

Table 3. 4 Comparison of DGM- Using Modified Process Scoping Characteristics

	CS	DES	VM	MPM	PMTS	VDG
Time Required to develop initial data generation model (Weeks/Months)	M	M	M	W	M	W
Time required to carry out data generation experimentation (Weeks/Months)	W	M	M	W	M	M
Time required to collect the generated data (Hours/Days/Weeks)	H	D	H	H	W	W
Necessity to identify process or method experts for model development (Yes/No)	Y	Y	Y	Y	Y	Y

Table 3.5 shows the capability of the alternative data generation methods for generating data for different development stages of the product life cycle, and for a variety of production types and volumes. Referring to the aims of the research, Section 2.5, there is a need to identify a data generation method, which can overcome the lack of data, for CMD, at the concept design stage of a process or product. Therefore from Table 3.5, it was found

that only CS and VM are suitable methods to be used at concept design of product life cycle. PMTS is the most suitable for those tasks whose cycle times are small. DES is suitable to provide data where the effect of system delays, queuing periods and inventory are required. However, VM is the most ideal method for data generation for application where there is little or no historical data available, i.e. at the concept design stage (Alabastro, 1995; Williams, 107; Barrett 1997; Hollocks 1995; Shin 2000) and for low volume, one off projects and job shop environments.

Table 3. 5: Comparison of DGM- Product Life Cycle and Production Volume & Type

		CS	DES	VM	MPM	PMTS	VDG
Development state of process	Concept	√		√			
	Detail Design			√		√	
	Prototype			√	√	√	√
	Commercial		√	√	√	√	√
	New Product/Process			√		√	
	Modified Product/Process	√		√		√	
Production Volume	Low Volume/One Off	N/C	√	√			
	Medium Volume	N/C	√	√	√	√	√
	High Volume	N/C	√	√	√	√	√
Production Type	Project	N/C		√			
	Job Shop	N/C	√	√			√
	Batch	N/C	√	√	√	√	√
	Flow	N/C		√	√	√	√
	Continuous	√	√	√	√		
N/C –Not taken into consideration when developing and using the method							

The ability of CS and MPM to generate process time data depends on the form of the models used. In this respect it is normally necessary to include rate based variables within the models that are normally measured in terms of cost per quantity. The remaining methods DES, VM, PTMS and VDG are time based modelling methodologies, i.e. they model the behaviour of a process over time and hence time-based data is inherent.

DES requires input of the basic process times for operations. It is suitable for generating data concerning the indirect times that arises within work systems, e.g. such activities as "waiting in queues" or the "idle time of equipment". DES can identify the effect of operational dynamics but not the time taken to perform activities at a process level which are normally associated with the generation of specific product features using specific process features.

However, VM is normally a real time based simulation process in which the time taken for an individual activity is recorded by the VM data collection system. VM models are capable of generating output data at detailed levels including individual tool movements (Lee 2001). Individual process tasks can be identified and can be validated visually using both VM and DES. In order to determine if VM represents a suitable data generation method it is necessary to evaluate the method with the remaining cost modelling characteristics identified by Delgado, Stockton (2002), i.e. this evaluation is shown in Table 3.6.

Table 3. 6: Comparison of DGM- Level of Data Details

		CS	DES	VM	MPM	PTMS	VDG
Resource able to be estimated	Direct Material Cost						
	Indirect Material Cost						
	Direct Equipment Cost						
	Indirect Equipment Cost						
	Direct Labour Cost						
	Indirect Labour Cost						
	Direct Process Time	√	√	√	√	√	√
	Indirect Process Time		√			√	√
Product Feature levels	Product Level		√	√			√
	Component Level		√	√			√
	Component Feature Level			√		√	√
Process Feature levels	Machine Level		√	√	√		√
	Machine Assembly Level	√		√			√
	Machine Sub-Assembly Level			√			√
Process Activity levels	Process Level			√	√		√
	Process Operation Level		√	√	√	√	√
	Operational Activity Level	√		√		√	√

From Table 3.6, VM and VDG are the only methods capable of generating data different levels of product and process details. However, VDG method is not capable of using at the concept stage of product and process development. Finally it can be concluded that based on the analysis of Table 3.4, 3.5 and 3.6, VM is the one of the most suitable method for generating data where there is lack of historical data and process is in its conceptual stage of design.

3.5 Data Analysis Techniques

There are a variety of methods available for analysing data in order to identify its underlying cost estimating relationships (CERs). This section compares the traditional method of regression analysis with the more advanced methods, that use neural networks, fuzzy logic and data mining algorithms to model complex relationships.

3.5.1 Regression Analysis

Regression analysis is the traditional data analysis method used for developing CERs (Smith 1997). According to Mackenzie (2003) the least-squares method is the most widely used of the "best-fit" approaches within the range of regression-based techniques. It identifies the relationships between a dependent variable and one or more independent predictor variables (Roy 2001a). In terms of cost modelling 'cost' represents the dependent variable and the factors affecting these costs represent the independent variables. Regression analysis techniques include both single and multi-variable types based on the number of independent variables being examined. Cochran (1976) has used multi-variable linear regression approaches in which, product cost was the dependent variable and the product parameters the independent variables.

Estimating models developed using regression techniques are normally able to provide unbiased and efficient estimates providing the relationships between the dependent and independent variables are linear. Regression techniques cannot distinguish whether the relationships found between variables are in fact causal. Hence, independent variables must be chosen with care. The CERs obtained using regression analysis is normally tested for "best fit" using the coefficient of determination, i.e. R^2. The nearer the value of R^2 is to 1 then the more statistically sound is the relationship between the dependent variable and its predictor variables. According to Dysert (1999) regression analysis is easy to use but can be a time consuming process since it often involves trial and error experiments to determine the most suitable combination of predictor variables. This tends to suggest that predictor variables are chosen on the basis of maximising estimating accuracy rather than whether they have a true causal effect on the independent variable. In addition, Dysert (1999) emphasised that most CERs are non-linear in nature, and therefore, users may need to experiment both with linear and non-linear variable combinations.

According to Melin (1994) the accuracy obtained from the use of regression analysis is based on the accuracy of the input data from which the CER is derived. According to Stewart (1995) the availability of only a small number of data points from which to derive CERs may reduce the number of independent variables that can be included in the CER.

3.5.2 Neural Networks

The increasing intensity of cost-based competition motivated research aimed at developing more advanced tools such as artificial neural networks (ANNs) for supporting cost estimating analysis (Shtub 1993), (Smith 1997). ANNs are trained using a learning function, the most common of which is the "error back propagation algorithm" (Bode 2000). Training enables the ANN to model the functional relationship between one or more dependent variables and their predictor variables. According to Denton (1996) the successful application of neural networks in a wide variety of areas can be attributed to the use of feed forward structures in which the data is trained from external data sources. This research criticised the use of back propagation algorithm due to their slow training speeds and need to specify learning rates, which themselves were found to be critical to successful training.

Bode (1995) proposed a 3-layer perceptron in order to identify the relationships between cost drivers and the cost of a product. Experiments were carried out by artificially generating the data for the purpose of ANN training. The levels of accuracies obtained from the resulting experiments were found to be within the range of 40% to 80%. It was also identified that with increasing amounts of training data the accuracies of models also increased.

According to Bode (2000) ANNs perform better for cost estimation when there are a small number of cost drivers, when there is sufficient data to support the training process and when there is sufficient information about the process, i.e. in terms of the influence of individual variables. Graham (2004) found that ANNs are suitable for solving problems that are difficult to measure and where process parameters involve a high degree of non-linearity in their relationship to cost.

The effectiveness of the regression technique for developing CERs has been compared with ANN techniques. Sonmez (2005) developed conceptual cost models for building projects using regression analysis and neural networks. This work concluded that regression analysis unlike ANN's cannot deal with modelling situations where both linear and non-linear relationships exist within the same model.

Kim (2004) also compared the use of regression analysis with artificial neural networks and case base reasoning and identified that regression analysis could not assist the cost model developer in selecting those variables that produced cost models that best fit the available historical data. He suggested that there was a need for the process variables, to be used within the analysis, to be reviewed for validity in advance of their use. In addition, when compared with artificial neural network and case base reasoning the number of variables that could be included in the model was severely restricted. Wang (2000) also compared multiple linear regression with artificial neural networks and found that as "number of variables" increased the estimating accuracy also increased. However, this work also found that there was no significant increase in the estimating accuracy with increasing "number of data points".

Shtub (1999) compared simple linear regression analysis and ANNs for estimating the cost of steel pipe bending and compared the cost models developed using these data analysis techniques in terms of the accuracy of the results obtained. This comparison revealed that for the same number of variables and data points the model obtained using ANNs provides higher estimating accuracies than regression based models. Also, Sonmez (2004) states that except for the selection of an appropriate architecture for neural network models, the user does not need to put additional effort in deciding the nature of relationships between the variables. On the other hand, regression models use fewer model variables when compared to ANN for prediction, which may give better prediction if the relationship between the variables is of linear nature.

Zhang (1996) identified designed specific ANN architectures for particular cost modelling applications. According to Denton (1996) building ANNs with high levels of estimating accuracy can be problematic due to difficulties in selecting which input variables to use, and both the number of hidden layers and number of data points for training the ANN. Wang (2000) suggested that the successful application of neural networks for developing cost models was based on appropriate selection of all structural elements that make up an ANN. Here Wang (2000) developed a methodology, based on the use of Taguchi Orthogonal array, for selecting those structural elements of an ANN that provided the highest levels of estimating accuracy. The results revealed these ANN models outperformed those developed using regression analysis in terms of model prediction accuracy. However, the accuracies of multiple linear regression models were found to be adversely effected by increasing numbers of input variable types.

Research into the application of neural network for cost estimation by Setyawati (2003) also suggested that the number of data points and selection of input variables have a significant impact on model performance. In general the greater the number of data points available for training the better estimating accuracy of the ANN model developed. Contrary to this, Creese (1992) found that the prediction performance of neural network models when trained using randomly selected data is less than regression-based models. A major limitation, according to Zhang (1996) of the use of ANN models in the cost modelling domain is their 'black box' approach, i.e., the relationships, type and size, between individual predictor variables and the dependent variable are not explicitly known. This makes the resulting model and their subsequent cost estimates difficult to justify to those affected by its use. Hence, ANNs have not been widely accepted as a common practice in cost modeling (Idri, 2002).

3.5.3 Fuzzy Logic

Fuzzy logic according to Kay (2004) "is an extension of traditional Boolean logic designed to work with imprecise data, with the concept of partial truth. The traditional reasoning has 'yes' or 'no' values. Fuzzy logic can handle values in between such as "maybe" and "nearly"." To produce fuzzy rules linguistic expression is used to which truth-values are assigned. A fuzzy logic model comprises of several fuzzy sets of input and output variables. Each variable has a number of membership functions and the relationship between these can be explained in a qualitative manner, such as the cost of machining operations is high, average or low. The fuzzy logic approach to cost modeling can be applied when there is insufficient quantitative data available to develop CERs using traditional approaches such as regression analysis (Mason 1997) or where a high level of uncertainty is associated with data. Jahan-Shahi (1999) has applied a simple rule-based fuzzy approach to model the costs of the flat plate rolling process with input variables of plate thickness, plate size and labour skill level. Membership functions assigned to plate thickness were "tiny", "regular" and "thick"; those to plate size were small, medium and high and those to labour skills it was "poor", "good" and "excellent". With the help of a rule based fuzzy approach the subjective cost drivers transformed into quantifiable variables that were able to generate reliable and comprehensive cost estimates. This is due to the ability of fuzzy systems to implement subjective concepts and imprecise information to estimate processes with insufficient information or data. According to Shehab (2001) and Graham (2004) the basic steps involved in building a fuzzy logic system are:

(i) fuzzification of inputs,

(ii) fuzzy inference based on a defined set of rules, and

(iii) defuzzification of the inferred fuzzy values.

This ability to cope with qualitative and uncertain information makes the fuzzy logic approach particularly suitable for developing models that can estimate costs at the early stages of product design and manufacturing (Shehab 2002; Jahan-Shahi 1999; Zhao 2006).

Further application of the fuzzy logic modeling approach to develop CER was undertaken by Ping (1996) who developed a cost estimation system based on the use of a distributed multi-agent system. The model is based on the use of fuzzy classification of cost estimation methods and dynamic optimization of model structure using multi-agents. In this model the input data provided to the system is distributed to the multi-agents in order to make the agents 'learn' from the previous estimates.

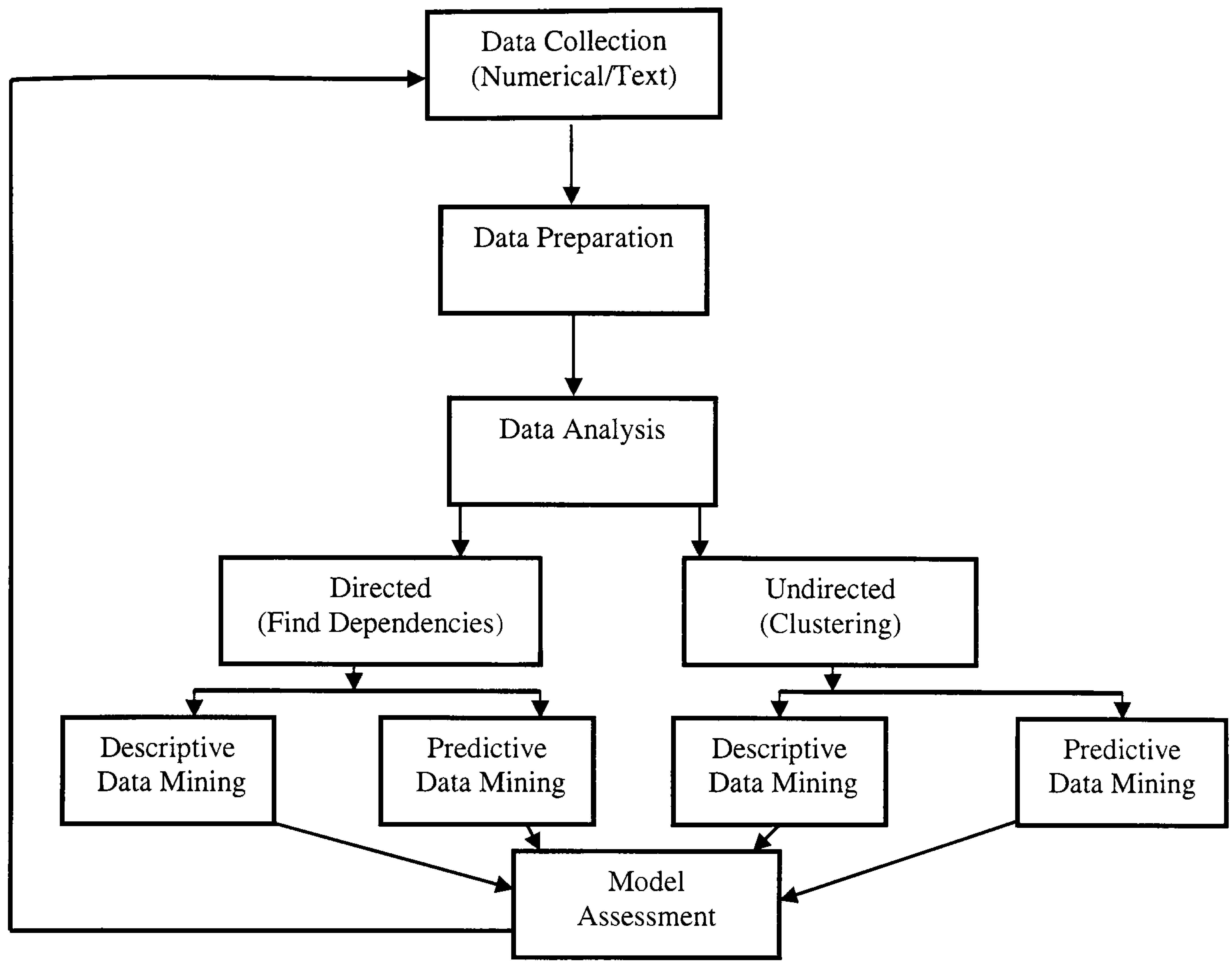
Zhao (2006) presented a cost estimation model for design and manufacturing of composite structures again for use during the early design stage of product development. The cost estimation model was based on the use of the Fuzzy Multi-Attribute Utility, (FMAUT), technique developed by integrating Multi-Attribute Utility Theory (MAUT) and fuzzy logic. MAUT assigns utility values to process cost drivers based on expert opinions and the fuzzy logic is used to estimate the cost of the process based on these assigned utility values. This enables both historical data and expert opinion to be combined to estimate costs. However, the accuracy of the models still strongly reflects the expert's level of knowledge to estimate the process and the degree of bias inherent in the expert's opinions.

Through the use of a fuzzy logic approach, the limitations of the traditional techniques, i.e., where only subjective and/or imprecise data is available, can be overcome. It is still difficult however to maintain degree of significance due to extraction of knowledge by logic (Mitra 2002).

3.5.4 Data Mining

Data mining is the process of 'discovering' the data relationships that may exist within large amounts of data. As with other traditional data analysis techniques such as multiple linear regression, data mining techniques involve the use of algorithms such as ARNAVAC, SKAT algorithm, GMDH algorithm, to determine relationships between a dependent variable and its predictor variables (Koonce 1997; Hosking 1997; Read 1999). However, according to Madhisetty (2004) data mining has greater efficiency when compared with these traditional data analysis techniques. Figure 3.1 illustrates the basic tasks involved in data mining process.

Figure 3. 1: Data Mining Process



Initially, the data mining process is categorised as directed and undirected, directed data mining is the one in which the dependent variable is known, and the user knows what to predict. In undirected data mining the dependent variable is not know and the user has to analyse and explore the data in order to establish relationships among the variables. Both of these, directed and undirected data mining, can be carried out using descriptive and predictive algorithms based on the type of data. Both types use a variety of statistical methods and algorithms to undertake the data mining process, i.e. Table 3.5.

Table 3.5 Data mining Methods

Predictive Data mining	Descriptive Data mining
Linear Regression	Taxonomy Categorizer
Find Laws	Text Analysis
Memory Based Reasoning	Text Categorisation
PolyNet Predictor	Text Derepeater
	Text OLAP
	Link Analysis

Descriptive methods, which are referred to as unsupervised methods, group data into interpretable patterns. Most commonly used descriptive data mining methods make use of cluster analysis, Kohonen maps, association rules, log-linear models and graphical models (Giudici 2003).

Predictive methods, which are referred to as supervised methods, identify the quantitative relationship between predictor and dependent variables. Most commonly used predictive data mining methods are regression analysis, classification methods, example-based methods, probabilistic graphic dependency models, relational learning models and dependency modelling (Fayyad 1996).

The knowledge extracted from data mining has been used in a wide variety of decision-making areas and at various levels of decision-making. It has been applied within such areas as product marketing for identifying buyer trends for particular products (Ling 1998), the Pharmaceutical industry for decision-making during compound selection in drug discovery (Balakin 2004), Health care diagnosis for disease modelling (Fazel 2003), the Retail Industry for analysing customer behaviour in terms of requirements inventory

(Haigang 2005), Financial institutions for fraud detection (Knobbe 1997) and manufacturing sectors for process control (Sadoyan 2006, Thamaraiselvi 2004, Kusiak 2000). Also, in manufacturing, data mining has been used to discover new rules for quality control, process control, defect prevention, and safety and performance evaluation (Kasravi 1997). Loss (2005) presented the use of data mining for exploring the hidden patterns and knowledge from virtual enterprise data. This was carried out using data mining algorithms, i.e. Clustering and K-Means, to facilitate the decision-making process. In this way data mining process is normally applicable to a wide range of applications (Giudici 2003).

A number of commercial data mining systems are available of which Polyanalyst (Megaputerintelligence.com) is a typical example and has been used within the current research. PolyAnalyst undertakes the data mining process in two basic stages, i.e.:

i) it identifies a sub-set of the total dataset within which meaningful relationships exist between nominated dependent variables and predictor variables, and
ii) it then quantifies the relationships within this subset between the predictor variables and the dependent variable.

PolyAnalyst is, therefore, a multi-strategy data mining system that implements a broad variety of mutually complementary methods, e.g. Discriminate and Find dependencies that are able to automate the data analysis process.

3.5.4.1 Find Dependencies

The Find Dependencies algorithm is used to find the relationships that exist between the target dependent variable and its independent predictor variables without actually establishing the specific quantitative form of these relationships, i.e. it basically identifies how strong these dependencies are. The Find Dependencies algorithm is, therefore, primarily used in the pre-processing stage of the data mining process, which is aimed at providing processed data to the secondary stages of the data mining that deal with establishing the precise quantitative relationships. The Find Dependencies algorithm identifies a sub-set of data with in the complete dataset, which obey specific dependencies and removes data that does not obey those found specific dependencies. Hence, FD facilitates the identification and removal of flaws, outliers and errors in the data sets. Pre-processing using the Find Dependencies algorithm enables the determination of the variables and data sub-set that are of most importance to the subsequent data analysis process.

3.5.4.2 Stepwise Linear Regression

PolyAnalyst's stepwise linear regression automatically produces the best linear prediction rule for a particular dataset and the variables it contains. The method is based on a machine learning algorithm that performs the multi-parametric linear regression search on a variable number of parameters, with automated selection of the most influencing independent variables. Also provided are the values of R-square, standard deviation and standard error which indicate the estimating accuracy of the resulting model.

3.5.4.3 Find Laws

The Find Laws algorithm is used to identify nonlinear relationships between a dependent variable and its independent variables and to generate predictive rules for datasets. Find Laws is based on the Symbolic Knowledge Acquisition Technology (SKAT), which is able to explore the functional dependencies among the variables in the dataset and represent the discovered knowledge or rule in the form of mathematical equations that may include rational polynomials, relational operators and conditional blocks. The ability of Find Laws to automatically build a wide variety of mathematical constructions, including complex nonlinear algebraic expressions and functions, makes it an efficient knowledge discovery tool. However, Find Laws is resource-intensive and takes a great deal of time to complete its function. As a result, Find Laws is often used in the final stages of data mining to present a human-readable rule explaining the analysis.

3.5.4.4 PolyNet Predictor

PolyAnalyst's PolyNet Predictor makes use of artificial neural networks to generate a network of nodes in the form of mathematical expressions. If the resulting network nodes are simple and readable then they can be presented as actual mathematical expressions. Normally where there are large numbers of data points and variables the neural network process only operates in its "black box" mode. This means that neural network can generate valid predictions but in most of the cases it does not provide the details of how these predictions are arrived at.

The selection of a particular data-mining algorithm is primarily based on the type of data present. In addition, there is often no specific data mining methods that can provide a best solution, more often combinations of different methods may produce improved results (Buchner, 1997; Hosking 1997). Therefore, the choice of which algorithm to use is a subjective decision based on a user's prior knowledge and the type of problem that needs addressing.

The increasing use of data mining is due to the fast computational analysis speed possible, which overcomes a major drawback in the use of traditional statistical techniques. However, advanced data mining methods have not yet fully replaced the traditional proven statistical techniques such as the generalised linear models (Kolyshkina 2004). In this respect Kolyshkina attempted to enhance the effectiveness of the linear modelling approach by integrating this method with data mining tools in order to accomplish better accuracy in less time and with good levels of interpretability. The results obtained from this work proved that model building using this integrated approach was faster and higher levels of estimating accuracy could be achieved. Rocke (1998) confirmed this result, i.e. the integration of statistics and data mining provided methods for improved data analysis.

The accuracy and confidence in the results obtained from predictive data mining is critical to the validity of the methodology or model being developed. In an attempt to find ways for increasing confidence in the results obtained from data mining (Kusiak 2000; Lee and Morey 1994; Klein 1993) explored methods of improving robustness, i.e.:

a. including redundant data in the prediction algorithms, and
b. using independent methods to verify the same results.

Although data mining can identify previously unknown relationships between the variables, it cannot identify the significance of these relationships. To determine levels of significance, knowledge of the underlying phenomenon must be known. Further, limitation of this technique is that although it can identify the key relationships between variables, it does not necessarily identify the causal relationships (Seifert 2004).

3.6 Comparison of Data Analysis Techniques

From Section 3.5, key requirements have been extracted and used to provide a comparison, Table 3.7, of the alternative data analysis techniques. Table 3.7 indicates that all four of the data analysis techniques appear to have the ability to identify and prioritise the cost drivers and to develop cost estimating relationships from the identified set of cost drivers. Neural Networks, Fuzzy Logic and Data Mining seem to possess common characteristics in each such as the ability to deal with both small and large numbers of variables, small and large numbers of data points and both linear and non-linear relationships between variables. However, it is only the data mining process that has the capability to produce multiple cost output types which contain both linear and non-linear terms. Therefore, it is the Data Mining process that best meets the criteria listed in Table 3.7 for selection of a data analysis tool for the cost model development process.

Table 3. 7: Comparison of Data Analysis Techniques

	Linear Regression Analysis	**Neural Networks**	**Fuzzy Logic**	**Data Mining**
Fast and Interactive		√		√
Identification and Prioritisation of cost drivers	√	√	√	√
Numbers of data points required	M	M	M	F & M
Number of variables required	F	M	M	F & M
Types of Relationships	L	L & NL	L & NL	L & NL
Provide multiple outputs				√
Provide predictive model not Black box	√		√	√
F-Few, M-Many, L-Linear, NL-Nonlinear.				

Chapter 4 Experimental Design

4.1 Introduction

From Chapter 3 it was identified that Virtual Manufacturing is the most suitable method in terms of its ability to generate data at all level of process detail. It can also provide a means of validating process models visually. Similarly, Data Mining was found to be a suitable data analysis method for use within the cost model development process at the early design stage. Chapter 4 now examines the application of these tools in detail. The objective of this chapter is to develop a series of experimental trials by which the validity of the following can be identified i.e.:

i) the suitability of VM as a data generation method for cost modelling of processes and / or products at their conceptual stage of development, and

ii) the suitability of data mining as a method for analysing data from a VM environment such that accurate estimating models may be developed.

In addition, this chapter will examine the effects of "the number of process variables" and "the amount of data" on the effectiveness of data mining techniques.

4.2 Manufacturing Processes for CMD

The manufacturing processes selected for the purpose of testing the VE-based data generation process were Vertical End Milling and Automated Spray Painting. These manufacturing processes were selected because they represented differing levels of process complexities and process types and contained:

a) a range of product features effecting process times, e.g. machining length, machining profile, depth of cuts, work piece materials, and work piece geometries, and

b) a range of process features affecting process times, eg. surface speeds, feed speeds, feeds per tooth and spindle speeds.

4.2.1 Vertical End Milling Process

The basic information required to build a model of a Vertical End Milling process was as follows:

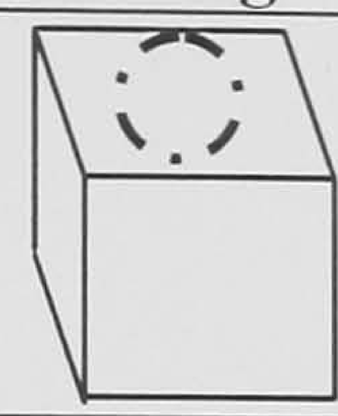
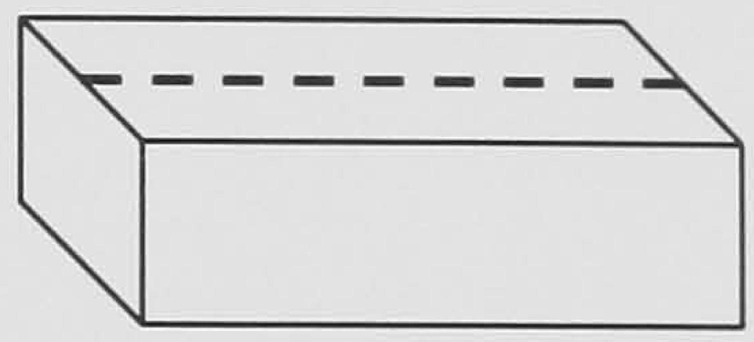
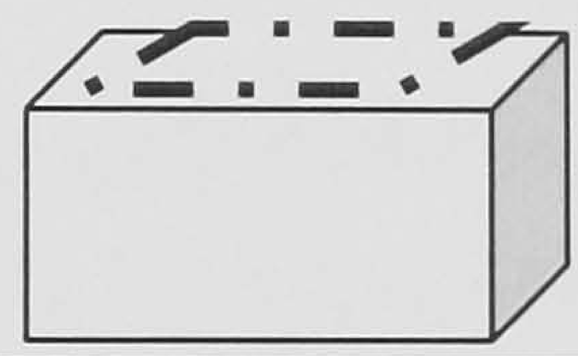
i) Process Description

This process consisted of the End Milling of slots which is a multi-point cutting process in which material is removed from a work piece by a rotating tool. Both the end and the periphery of the tool remove the material. The cutter rotates about an axis perpendicular to the surface.

ii) Product Characteristics

The process includes a range of product features, that center around the design geometry of the product and its material type. For the experiments three different material types were selected, i.e. cast iron, carbon steel and nickel base high temperature alloys. Similarly, three different machining profiles were selected, i.e. round, straight and square, the minimum and maximum machining length of each profile are listed in Table 4.1.

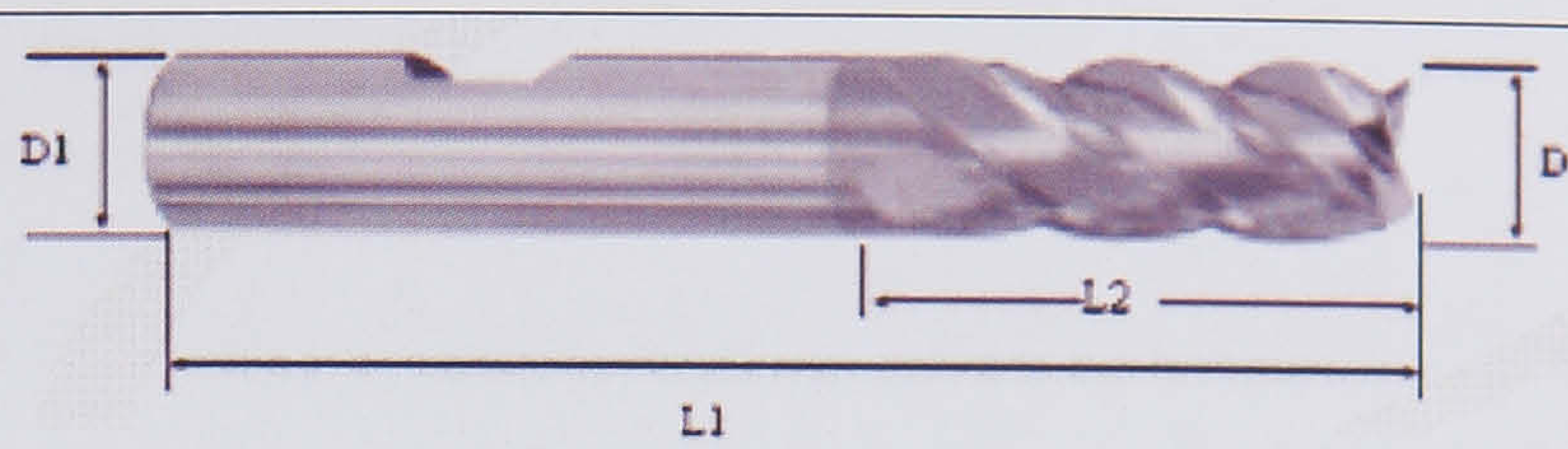
Table 4. 1:Range of Product Parameters Features

Product Feature	**Features selected**	**Product Drawing**
Material Type	Carbon Steel	
Machining Length	381 to 1524 mm	
Machining Profile	Round	
Material Type	Cast Iron	
Machining Length	762 to 2032 mm	
Machining Profile	Straight	
Material Type	Nickel Base High Temp Alloy	
Machining Length	115 to 680 mm	
Machining Profile	Square	

iii) Cutting Tool Description

Cutting tool descriptions are taken from the tool manufacturer's manual (Frasia, 2004). Table 4.2 lists the range of values selected for providing a suitable range of working conditions. Tools are described in terms of D (the diameter in mm of the cutting section), D_1 (the diameter in mm at opposite end being cut), L_1 (the full length in mm of the tool), L_2 (the tool cutting length in mm or flute portion) and z (the number of flutes in a tool).

Table 4. 2: Range of Cutting Tool Variables & Drawing

Tool	Units	Range Selected	Tool Drawing (Frasia, 2004)
D	mm	6.4 to 50.8	
D_1	mm	6.4 to 50.8	
L_1	mm	63.5 to 140.7	
L_2	mm	30.48 to 63.5	
z		3 to 5	

iv) Cutting Conditions

The range of cutting conditions selected is listed in Table 4.3. The range of values for these process variables are selected from the manufacturer's manual (Fraisa, 2004) and are based on the material type being cut. Depth of cut (Dc) is selected based on the diameter of cutting tool. Surface speed (Vc) and feed per tooth (F_t) data were again selected based on the material being machined. Using Vc and F_t the feed speed (Vf) and spindle rotations per minute (n) were calculated using appropriate formula (Creese, 1992).

Table 4. 3: Range of Milling Machining Parameters

Process Features	Units	Material Type		
		Cast-Iron	**Carbon Steel**	**High Temp Alloy**
Vc	m/min	76 to 106	152 to 304	19 to 31
F_t	mm	0.025 to 0.127	0.025 to 0.685	0.025 to 0.127
Vf	mm/min	582 to 2037	292 to 8552	35 to 762
n	rpm	470 to 10696	3820 to 30560	115 to 3056
Dc	mm	7.7 to 19	15.2 to 38.1	1.9 to 7.6
Lc	mm	762 to 2032	381 to 1524	115 to 680

v) **Resources**

The resources used to build the VM milling process models were CNC machines of the following makes, i.e. HAAS, MANDELLI Machine and EMCO VMC100.

4.2.2 Vertical End Milling Process Model

The virtual process models for Vertical End Milling were developed using the process data identified in Section 4.2.1. Three models were developed of varying complexities in terms of cutting profiles, number of axes degrees of freedom, depths of cut, tool types and material types. The objective of these models were to:

i) generate heterogeneous data in order to test the ability of data mining to identify the hidden relationships among the selected product features and process activities,

ii) identify the time required to build these models, generate data and finally to collect data, i.e. the key tasks of the modified process scoping task shown in Table 3.4, and

iii) identify the effect of product and process features on the milling machining times by using three different cutting profiles i.e., round, straight and square.

Figures 4.1, 4.2 and 4.3 illustrate the three milling process models that were developed using the Delmia D5R13 VNC Virtual Manufacturing software package (Delmia, 2004).

Figure 4. 1:Virtual Milling Model-1 (Straight Cut) using the HAAS CNC Machine

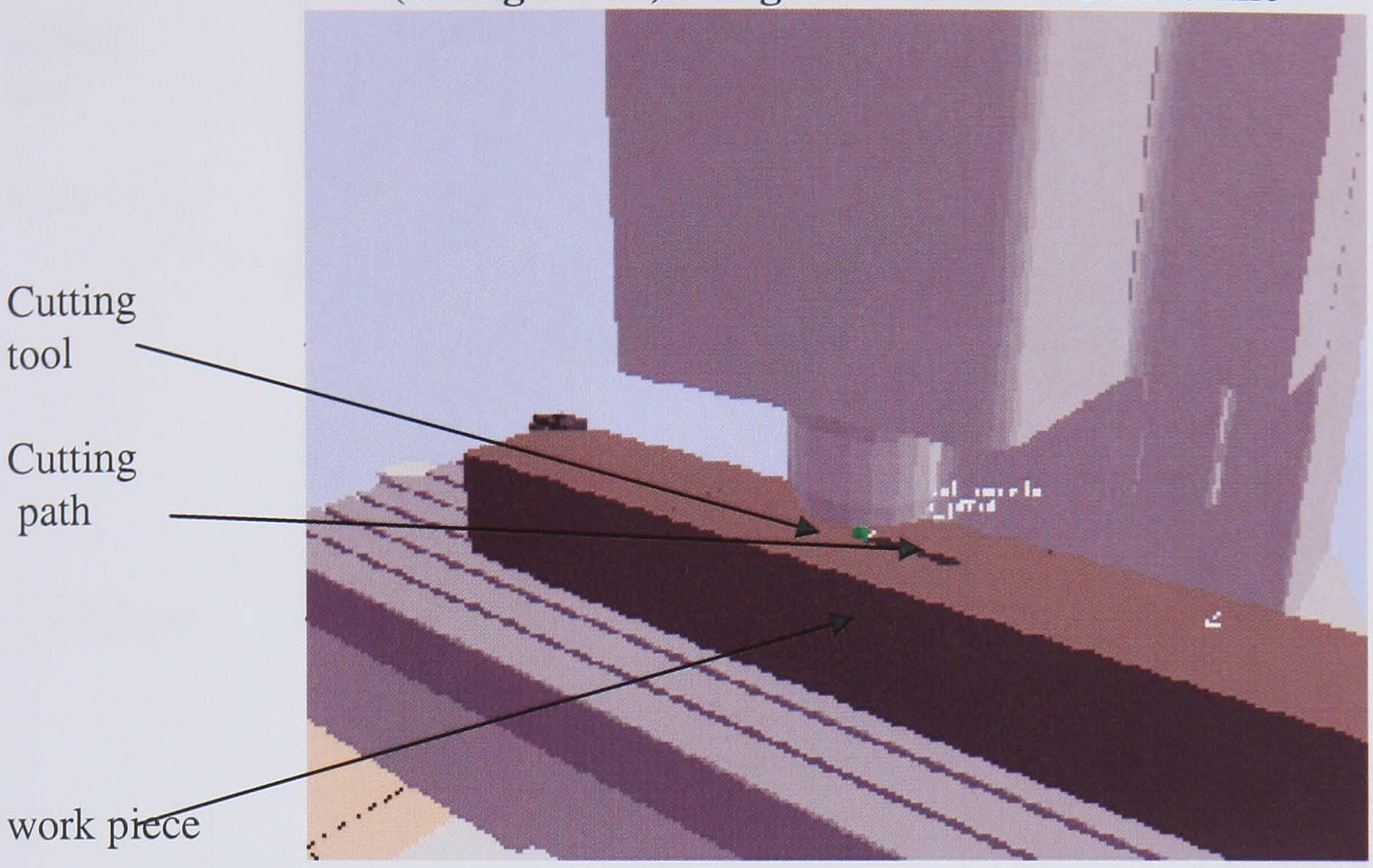


Figure 4. 2: Virtual End Milling Model-2 (Square Cut) using the MANDELLI machine

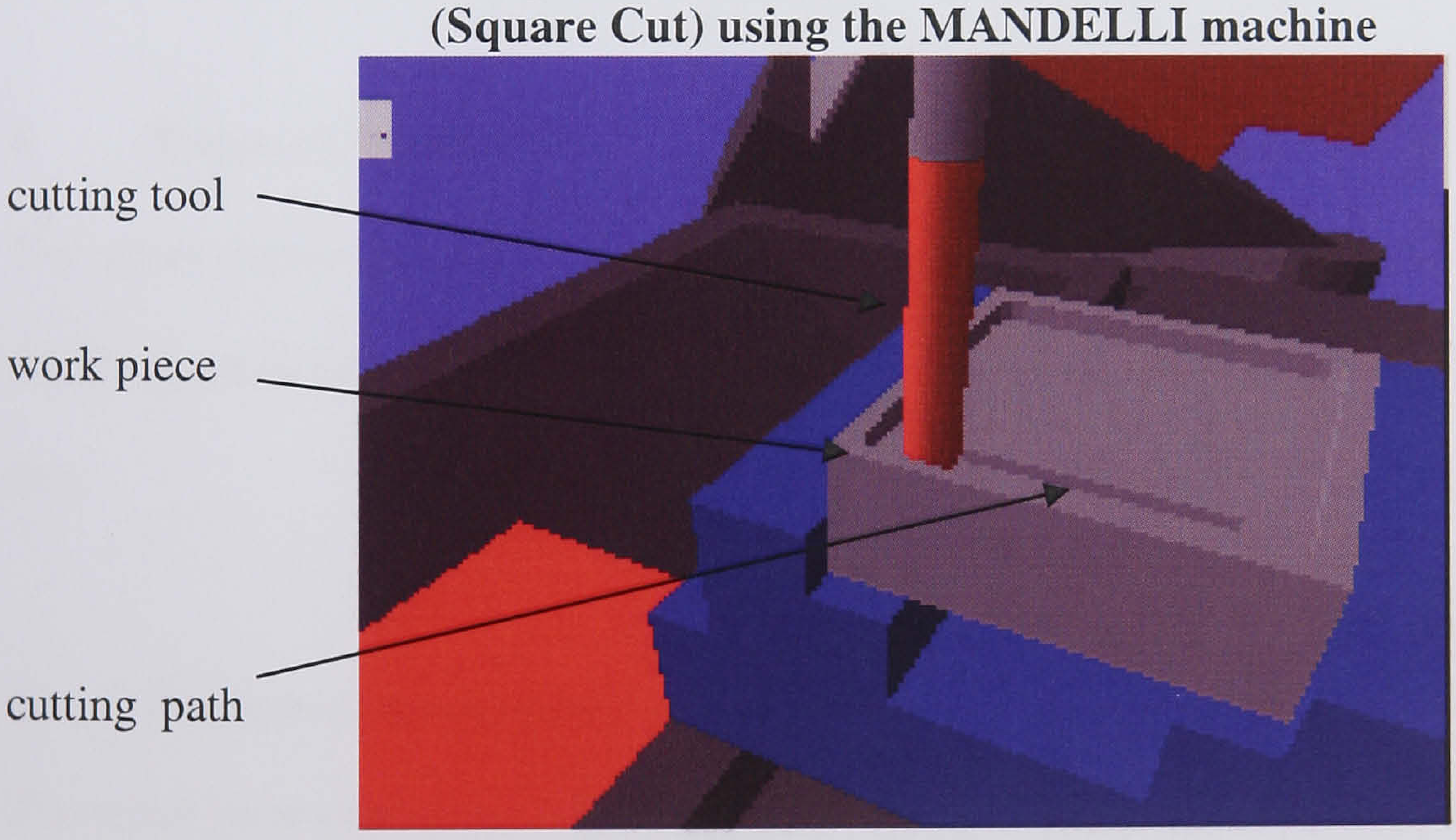
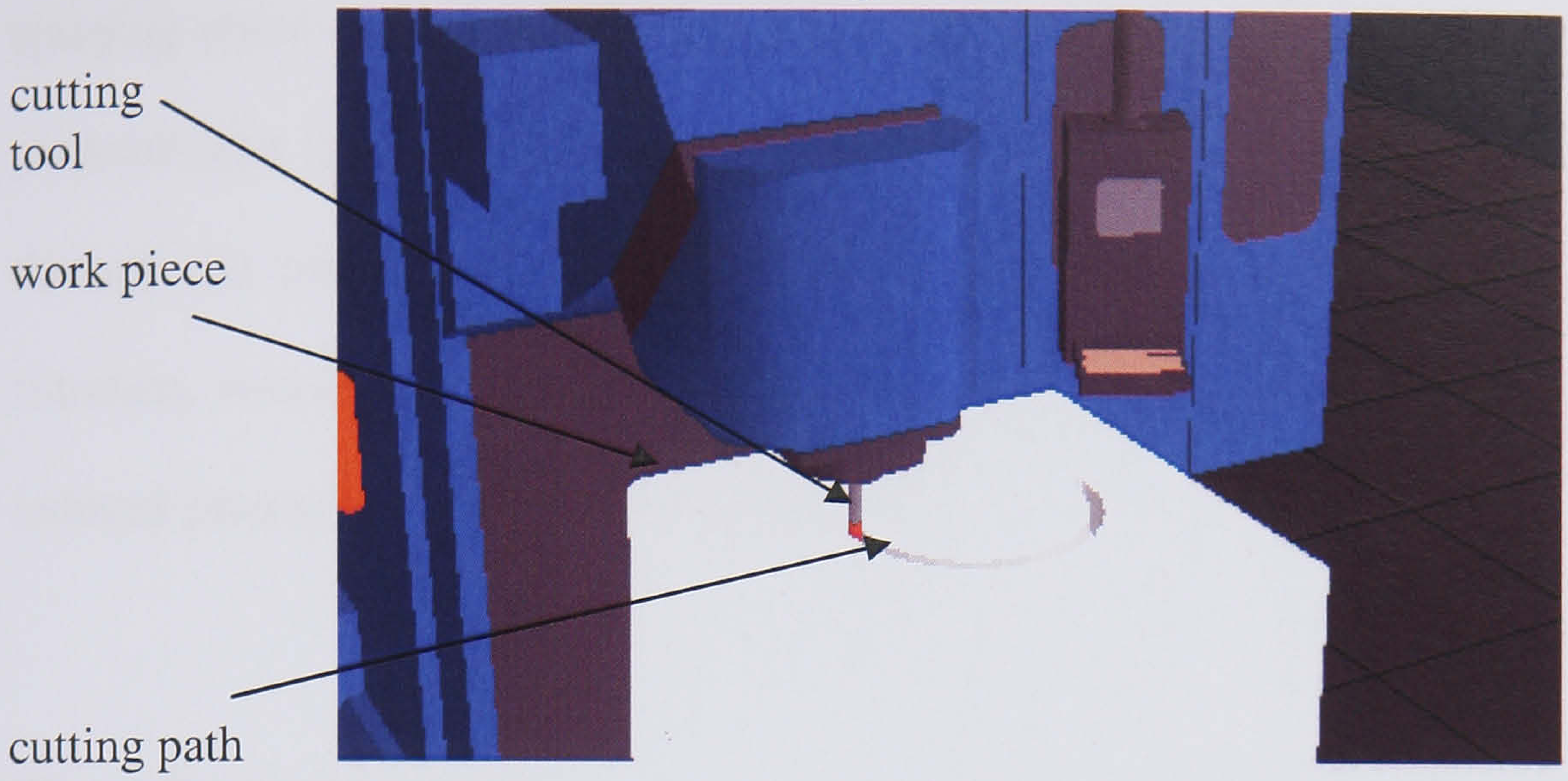


Figure 4. 3: Virtual End Milling Model-3 (Round Cut) using the EMCO VMC 100



4.2.3 Automated Spray Painting Process

The basic information required to build the spray painting process model was as follows:

i) Process Description

The spray painting process chosen was a fully automated process using a 6-axis robot which spray paints a target work piece which is fixed to a continuously moving conveyor belt.

ii) Process characteristics

The robot paint line is capable of automatic application of spray paint coating. The painting booth is equipped with a paint robot, an over hanging conveyor, a fixture and work pieces. The robot with a spray gun attached can be moved in different planes of a co-ordinate system based on the geometry of the part. Spray painting points are defined based on the geometry of the product and the painting requirements. The spray gun moves along the spraying points completing the programmed path. The direction of paint spray is always perpendicular to the surface being covered. A conveyor system transports the work piece through the paint booth to the curing area. Parameters such as relative humidity, air filtration, viscosity of paint, and spraying booth temperature are established using standard industry practice (Klein 2002, Asakawa 1997).

iii) Product Features

The target work pieces chosen for spray painting were of different geometrically defined shapes and of varying complexities, since, it is necessary when costing products to identify the effect of product features on the process cycle times and hence the process cost. The target work pieces of widely differing shapes, used for automated paint spray were a connecting rod, a car bumper, a truck door and an airplane wing. The 'surface area' of the target work pieces to be painted, measured here in terms of the distance traveled by the paint gun to paint spray these areas, is a key product features.

iv) Spray Gun Specifications

The Transfer Efficiency (TE) of a paint gun represents the amount of material that adheres to the target work piece as compared to the amount of material that was sprayed through the spray gun towards the work piece. According to Klein (2002) a typical high volume low pressure spray gun has a TE of 65%. A further factor is the spray gun distance from the target surface, which contributes towards the amount of paint that evaporates before reaching the target surface. Table 4.4 lists the range of parameter values for the paint gun specification chosen for the experiments.

Table 4. 4: Range of Paint Gun Parameters

Parameters	**Range**
Paint Flow Rate (cc/min)	200 to 750
Paint gun speed (mm/sec)	300 to 750
Distance from Workpiece (mm)	175 to 350
Paint Gun Range(mm)	300 to 550
Solids by Volume (%)	50 to 80
Gun Transfer Efficiency (%)	75 to 90

Here *Fluid flow rate* represents the rate at which paint is sprayed on target by the paint gun, *paint gun speed* represents the speed of the paint gun, *distance from workpiece distance* represents the distance between work piece and paint gun, *paint gun range* represents the maximum distance that a gun spray can cover and *solid by volume* (%) represents the percentage of total paint volume that consists of paint solids.

The paint deposition profile in the X and Y axes is termed the *paint brush specification.* Figure 4.4 shows the paint deposition profile that is perpendicular to the direction of the paint gun's motion profile.

Figure 4.4: Spray Paint Gun Profile

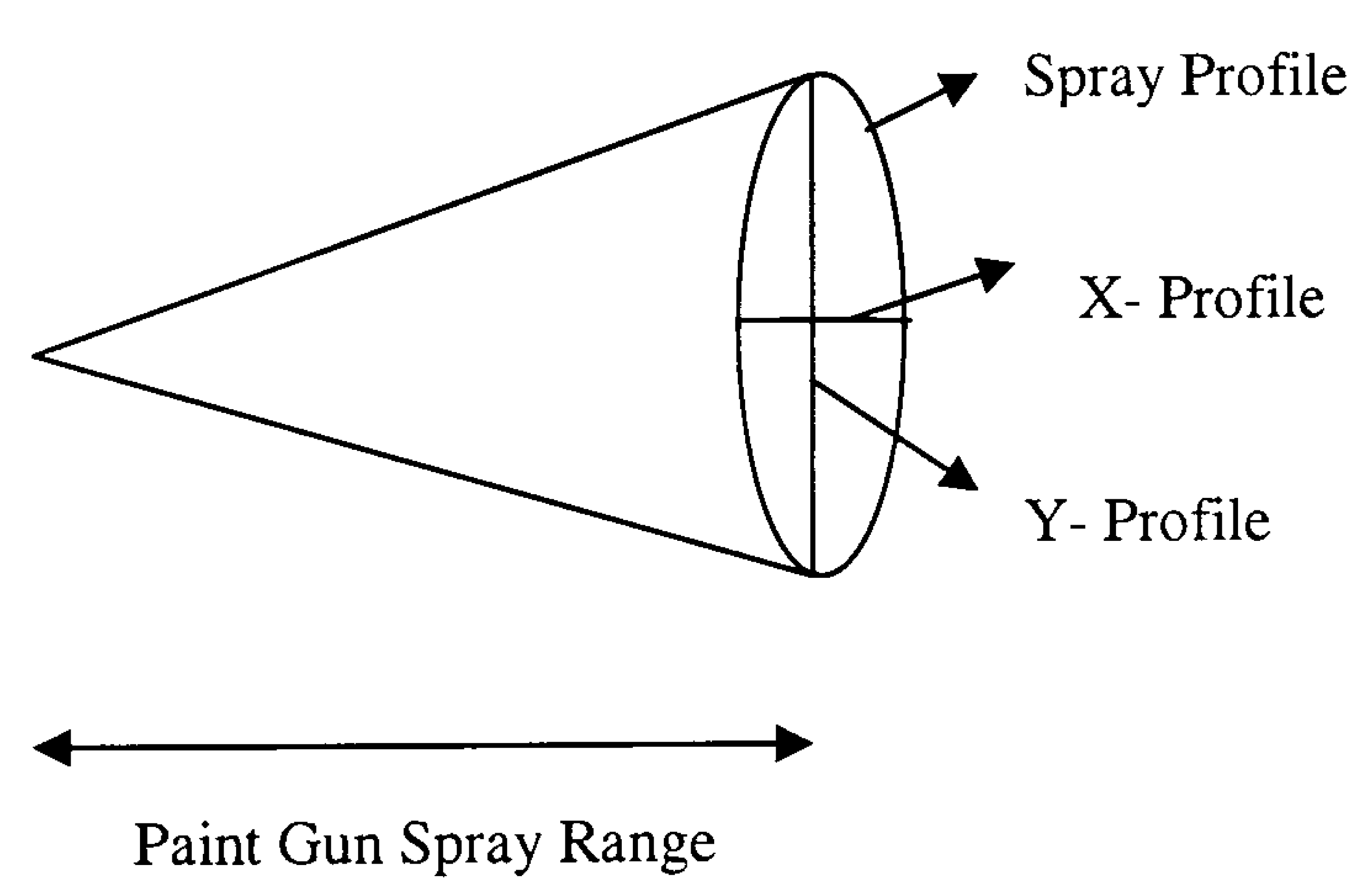


Table 4.5 lists the range of points on the X and Y coordinate systems of the Spray Gun Profile chosen for the validation experiments. In addition, this table also lists the paint thickness applied.

Table 4. 5: Range of Paint Gun Profiles

X-Profile	**Position (mm)**	**Paint Thickness (micron)**
Point 1	-150, -50	0
Point 2	-75, -20	35 to 105
Point 3	75, 20	35 to 105
Point 4	150,50	0
Y-Profile	**Position (mm)**	**Paint Thickness (micron)**
Point 1	-75, -20	0
Point 2	-150, -50	35 to 105
Point 3	75, 20	35 to 105
Point 4	75, 20	0

v) Resources

The resources required for an automated painting process are, 6-axis programmable robot, paint spray gun and overhead conveyor.

4.2.4 Automated Spray Painting Process Models

Four spray painting process models were developed using information provided in Section 4.2.3. Figures 4.4 to 4.7 illustrate these models, which were developed using the Delmia D5R13 UltraPaint virtual manufacturing software package (Delmia 2004). Here four different automated spray-painting models were used for the purpose of data generation and also to evaluate the time taken to develop these models, to generate data and finally to collect data. The over all description of the painting process carried out in these four models is described in Section 4.2.3 (ii). The time required in carrying out this task should be identified as this will help in evaluating the total time required to build cost models.

Figure 4. 4: Automated Paint Spraying Model-1 using Unimation Puma-761 Robot

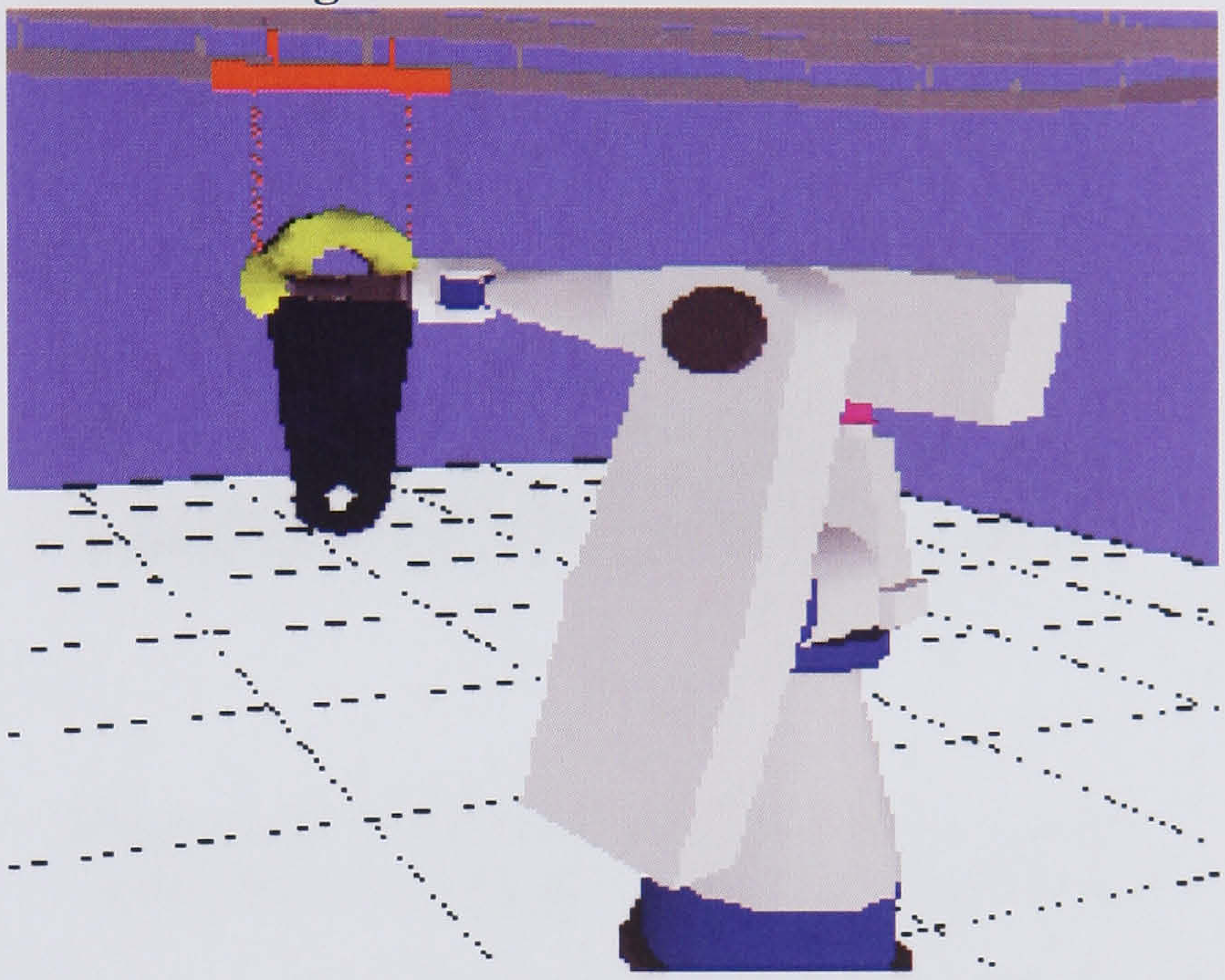


Figure 4. 5:Automated Paint Spraying Model-2 using ABB Fascia Robot (painting Car Bumper)

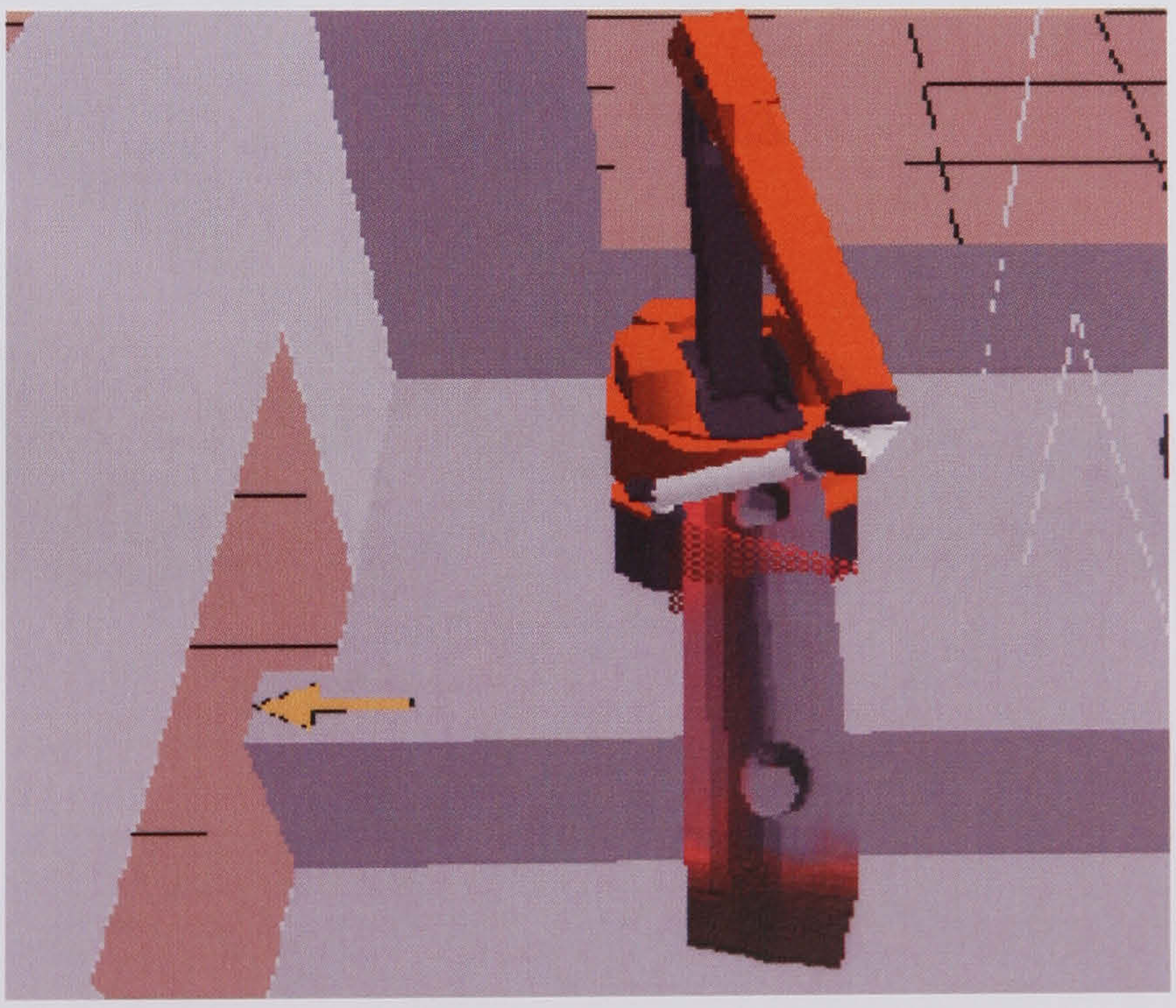


Figure 4. 6:Automated Paint Spraying Model-3 using ABB IRB 5403 Robot (Painting Truck Door)

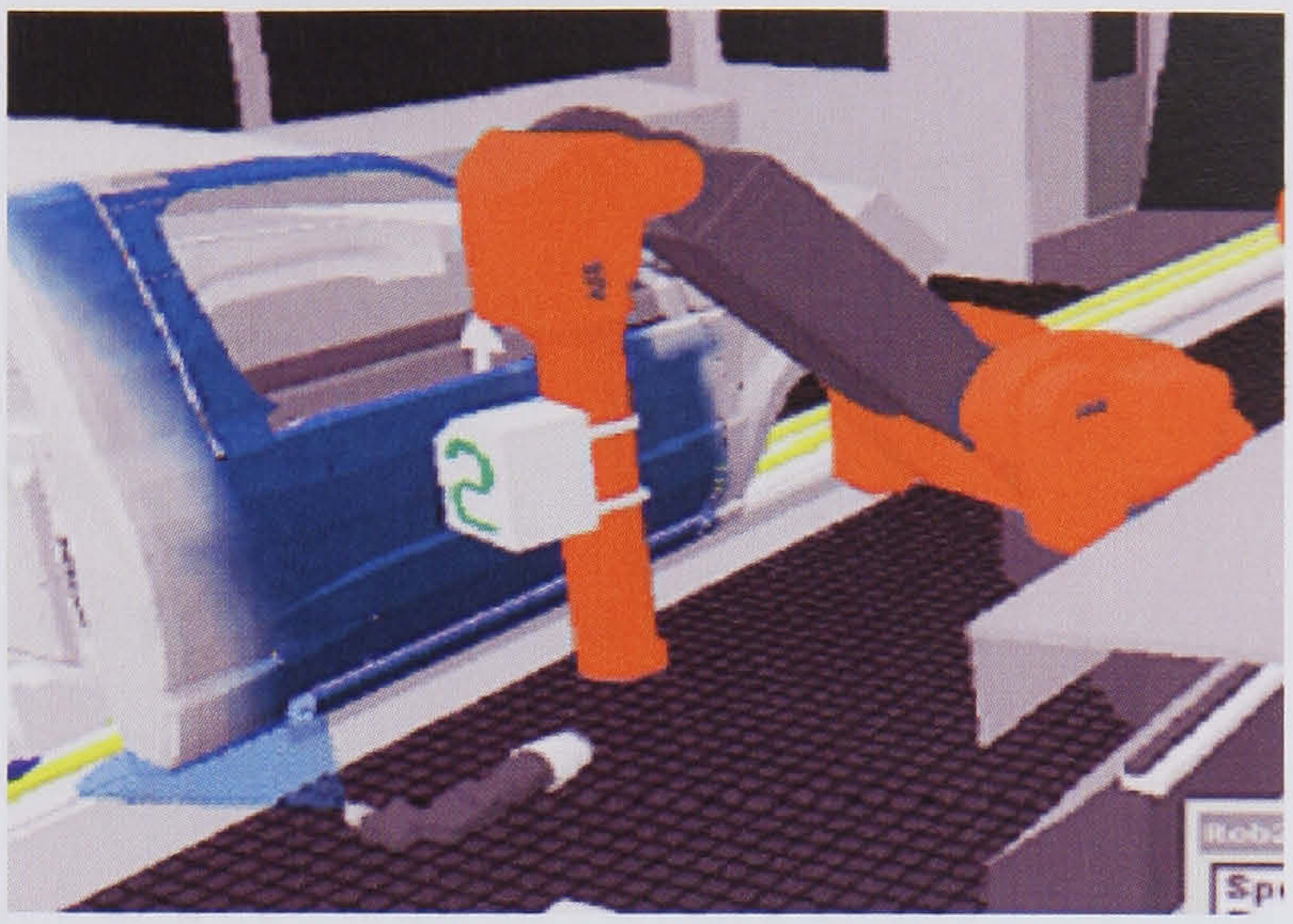
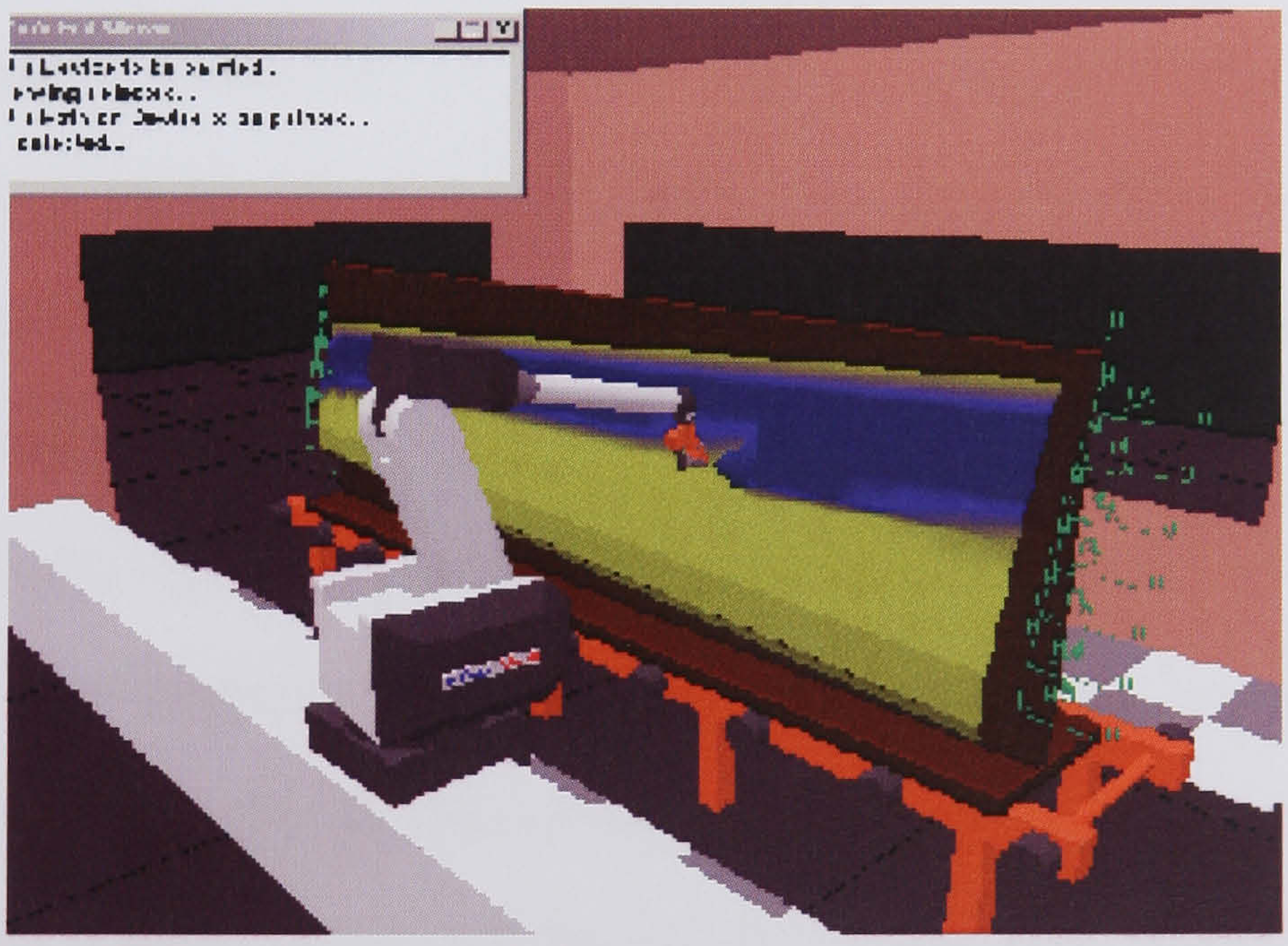


Figure 4. 7:Automated Paint Spraying Model-4 using GM Fanuc Rail Robot (Painting Wing)



4.3 Design of Experiments for Data Generation

Using the range of Vertical End Milling variables (Tables 4.1 to 4.3) and Automated Paint Spray variables (Tables 4.4 to 4.5) the Taguchi 'Design of Experiments' (DOE) methodology was used to developed experimental arrays in order to assist in the process of data generation. According to Antony (2004) in order to apply Taguchi's DOE approach, the following three steps are required:

i. identification of process variables to be used in the analysis,

ii. assignment of appropriate levels to these process variable values, and

iii. selection of an appropriate Orthogonal Array (OA).

The use of Taguchi Orthogonal Arrays (OA) are a resource efficient, cost effective and time saving methodology for designing experiments (Peace, 1993). According to Wang (2000) OA's facilitate the selection of a range of experiments capable of analysing large numbers of decision variables using the minimum number of experiments. Here, the use of the orthogonal array significantly reduces the number of experiments when compared to the use of a full factorial approach. A typical example of an orthogonal array is shown in Table 4.6. The first column in this table lists the experiment number, A, B, C then represent the different process design variables and the values *1, 2* and *3* in the columns represent the three values (i.e. levels) of each of the process variables. The selection of an OA is based on the number of individual variables to be examined and the number of levels chosen for each of these variables that need to be examined.

Table 4. 6: L9 (3^3) Orthogonal Array (MINITAB 2004)

Experiments	A	B	C
1	*1*	*1*	*1*
2	*1*	*2*	*2*
3	*1*	*3*	*3*
4	*2*	*1*	*2*
5	*2*	*2*	*3*
6	*2*	*3*	*1*
7	*3*	*1*	*3*
8	*3*	*2*	*1*
9	*3*	*3*	*2*

4.3.1 Application of the Taguchi DOE Methodology

In order to generate the data from which cost models could be developed experiments were designed using an appropriate OA. For each of the milling process variables three factor levels were selected, i.e. see Table 4.7. As described in Section 4.3, there are three material types and three cutting profiles considered in the milling process experiments. Hence an OA of configuration L27(5^3) for each of these profiles was designed. Table 4.8 shows the OA for the "square cutting" profile which is shown in Figure 4.2 representing Model-2. OA's for other two models and profiles are provided in Appendix 4.1, i.e. Tables 4.1A and 4.1B.

Table 4. 7: Milling Process OA Variables Levels

Process Variables	**Levels**		
	Level 1	Level 2	Level 3
Tool Diameter (D)	25.4	38.1	50.8
Depth of Cut (Dc)	15.2	25.4	38.1
Machining Length (Lc)	381	762	1524
Surface speed (Vc)	152	228	304
Number of Teeth (z)	4	4	5

The orthogonal array selected to design experiments for the Vertical End Milling process model-2 was the L27 which provides 5 variables and 3 levels per variable i.e. Table 4.8.

Table 4. 8: L27 (5^3) Orthogonal Array for Milling Experiment Model-2

	Lc	D	Dc	z	Vc
1	381	25.4	15.2	4	152
2	381	25.4	15.2	4	228
3	381	25.4	15.2	4	304
4	381	38.1	25.4	4	152
5	381	38.1	25.4	4	228
6	381	38.1	25.4	4	304
7	381	50.8	38.1	5	152
8	381	50.8	38.1	5	228
9	381	50.8	38.1	5	304
10	762	25.4	25.4	4	152
11	762	25.4	25.4	4	228
12	762	25.4	25.4	4	304
13	762	38.1	38.1	4	152
14	762	38.1	38.1	4	228
15	762	38.1	38.1	4	304
16	762	50.8	15.2	5	152
17	762	50.8	15.2	5	228
18	762	50.8	15.2	5	304
19	1524	25.4	38.1	4	152
20	1524	25.4	38.1	4	228
21	1524	25.4	38.1	4	304
22	1524	38.1	15.2	4	152
23	1524	38.1	15.2	4	228
24	1524	38.1	15.2	4	304
25	1524	50.8	25.4	5	152
26	1524	50.8	25.4	5	228
27	1524	50.8	25.4	5	304

Similarly, data generation experiments were designed for the automated paint spraying models. Table 4.13 shows the process variables selected for these experiments along with their values at three different levels. All process variable values except the path length is selected based on the based literature search and industry standards. Different values of path length for each model is calculated based on the different paths through which work piece can be painted, such as in horizontal direction, in vertical direction and in diagonal direction.

Table 4. 9: Spray Paint Variable Levels for Model-1

Process Variable	Levels		
	Level 1	Level 2	Level 3
Distance from Work piece(D_{wp}-mm)	175	225	275
Paint flow rate (P_{fr-} cc/min)	550	650	750
Paint Gun Speed (G_s-mm/sec)	600	700	800
Paint Gun range (G_{rg} -mm)	300	350	400
Path Length (P_{tl}-mm)	3500	6700	8300

Similarly, the OA's for automated spray painting were designed for each work piece using L27 (5^3) OA's, i.e. Table 4.14 represents the OA for Automated Paint Spraying Model-1 which is shown in Figure 4.4. OA's for other three models are provided in **Appendix 4.2, i.e. Tables 4.2A, 4.2B and 4.3C.**

Table 4. 10: L27 (5^3) Orthogonal Array for Model 1 Spray Paint Experiments

Exp.No	D_{wp}	P_{fr}	G_s	G_{rg}	P_{tl}
1	175	550	600	300	3500
2	175	550	600	300	6700
3	175	550	600	300	8300
4	175	650	700	350	3500
5	175	650	700	350	6700
6	175	650	700	350	8300
7	175	750	800	400	3500
8	175	750	800	400	6700
9	175	750	800	400	8300
10	225	550	700	400	3500
11	225	550	700	400	6700
12	225	550	700	400	8300
13	225	650	800	300	3500
14	225	650	800	300	6700
15	225	650	800	300	8300
16	225	750	600	350	3500
17	225	750	600	350	6700
18	225	750	600	350	8300
19	275	550	800	350	3500
20	275	550	800	350	6700
21	275	550	800	350	8300
22	275	650	600	400	3500
23	275	650	600	400	6700
24	275	650	600	400	8300
25	275	750	700	300	3500
26	275	750	700	300	6700
27	275	750	700	300	8300

4.4 Data Analysis Experiments

The analysis of the data generated from the Taguchi experiments was undertaken using a range of techniques involved in the Data Mining process. Experiments were carried out in order to compare the efficiency of individual algorithms at each stage of the data mining process. Table 4.12 lists the algorithms examined and the manufacturing processes used to examine their effectiveness. In addition to Milling and Painting process, Turning is also considered for the purpose building CERs and to evaluate the effectiveness of data mining process. The dataset for Turning was used from Wang (2000) consisting of 16 varaibles, which includes product, Turning process features and process activities identified by Boothroyds (2002). The Turning process considered here consists of turning machining tasks times which include roughening machining time, finish machining time, non-productive times e.g. set-up and effective area to be machined. The independent process variables involved in these tasks are categorized as product features, process features and process activities and are provided in Table 4.11.

Table 4. 11: List of Turning Process Variables

Product Features	
r_v	Proportion of initial volume
W	Weight of the work piece
Dm	Density of work material
A	Effective area to be machined
Ps	Specific cutting energy or unit power for the work material
Lr	Length /diameter ratio of work piece
Process Activities	
t_{sa}	Basic set-up time for machine
t_{ln}	Loading and unloading time
n_o	Number of operations
t_{pt}	Tool positioning time per operation
t_{sb}	Set-up time per tool
Process Features	
n_t	Number of tools
Ri	Proportion of material removed by internal machining
Re	Proportion of material removed by external machining
B_s	Batch size
Constants	
Rsg	A machinability factor

Table 4. 12: Experiments for comparing Data Mining Algorithms

Data Mining Process	Algorithms	Manufacturing Process
Pre processing	Find Dependence	Milling, Automated Spray Painting, Turning
Predictive Modelling	Stepwise Linear Regression	Milling, Automated Spray Painting, Turning
	Find Laws	Milling, Automated Spray Painting, Turning
	Polynet Predictor	Milling, Automated Spray Painting, Turning

4.5 Effect of Number of Variables and Number of Data Points on Model Accuracy

Data analysis experiments were carried out using the data generated to evaluate the effect of *number of variables* and *number of data points* on resulting model accuracies. In order to perform these tasks a similar approach was taken to Wang (2000) resulting in the experiments listed in Tables 4.13 and 4.14 being performed.

Table 4. 13: Automated Painting Experiments (Number of Data Points and Variables)

Experiments	Number of Variables	Variables	Number of Data Points
1-12	2	Gs, Ptl	300, 600, 900, 1200
13-24	4	Gs, Ptl, Pfr, Pns	300, 600, 900, 1200
24-36	7	Gs, Ptl, Pfr, Pns,Grg, Pnw,Dwp	300, 600, 900, 1200

Table 4. 14: Turning Experiments (Number of Data Points and Variables)

Experiments	Number of Variables	Variables	Number of Data Points
1-12	4	Bs, rv, ps, W	200, 350, 500, 750
13-24	8	Bs, rv, ps, W, dm, Rsg, ri, lr	200, 350, 500, 750
25-36	12	Bs, rv, ps, W, dm, Rsg, ri, lr, tln, tpt, tsb, nt	200, 350, 500, 750
37-48	16	Bs, rv, ps, W, dm, Rsg, ri, lr, tln, tpt, tsb, nt, re, n, tsa, no	200, 350, 500, 750

Chapter 5 Experimental Results

5.1 Analysis of Data Mining Algorithms

In order to identify and select the appropriate data mining algorithm for developing cost models, the experiments listed in Table 4.12 were carried out. The results from these experiments are listed in Tables 5.1 to 5.15 and illustrated in Figures 5.1 to 5.9.

5.1.1 Vertical End Milling

5.1.1.1 Data Pre Processing- Find Dependencies (FD)

Experiments were carried out using the dataset obtained from the Milling virtual process models described in Section 4.2.2. Initially, the dataset was explored using the Find Dependencies Algorithm and the outputs of these experiments are shown in Table 5.1.

Table 5. 1: Outputs from Find Dependencies Algorithm

FD Outputs	Definitions	Results
Size (No. of Data Points)	Represents the number of data points included in the CER analysis.	1679
IPV	Represents the list of all independent milling process variables selected for the purpose of identifying the key variables which can be included in the model.	D, Dc, Vc, n, Vf, Lc, Ft, Nt
DPV	Represents the dependent process variable selected in the CER analysis.	Tm
MIV	Represents the independent process variables selected by the FD algorithm for inclusion within the subsequent model identification stage.	Vf, Lc and Ft
Size (No. of Data Points)	Represents the set of data points where the dependent process variable distribution is maximally different from its overall distribution, and is considered to represent a subset of points revealing the dependence of the dependent process variable on independent process variables.	1433
P-value	Represents the probability that the detected relationship between dependent and independent process variables is merely a statistical fluctuation. The closer to zero the *P-value* is, the more significant is the reported dependence.	0

	The significance of the dependence i.e. P-value is calculated as P = $b(Nn_{FD}, N, A)$ Where $b(k, K, A)$ is a tail area probability of the binomial distribution where the number of trials equals K, and the event probability $A = \frac{1}{Nn_{cores}} \sum L_i$, where Li is the number of data points from the explored dataset whose target attribute values lay inside an i-th core region. Nn_{FD} denote the number of data points in an FD-subpopulation divided by N as n_{FD}.	

From Table 5.1, it shows that the Find Dependencies algorithm identifies a sub-set of data, within the complete dataset, which obey specific dependencies and removes data that is not related. The purpose of this is to identify if "exceptional" values existed in the data which if excluded by the FD algorithm may be helpful in development of the CER, i.e., lead to higher levels of estimating accuracy. Therefore, pre-processing using the FD algorithm enables the determination of the variables and data sub-set that are of most importance to the subsequent data analysis process.

CER Modelling

The next phase is the exploration of relationships among these variables identified by the FD algorithm, using the predictive data mining algorithms of Stepwise Linear Regression, Find Laws and PolyNet Predictor.

5.1.1.2 Stepwise Linear Regression (SLR)

The following model was developed using the Stepwise Linear Regression Algorithm, i.e.:

Equation (1);

$$T_m = (29.76 - 3.93D - 4.60D_c - 0.02V_c - 0.003n + 0.15 \times V_f - 545.34F_t) \quad \textbf{(1)}$$

Where:

T_m = Process time (min)

D = Milling tool diameter (mm)

D_c = Depth of cut (mm)

V_c = Surface speed of tool (m/min)

n = Spindle speed (rpm)

V_f = Feed rate (mm/min)

F_t = Feed per tooth (mm).

The statistical values shown in Table 5.2 provides indication of the estimating accuracy of Equation (2).

Table 5. 2: Estimating Accuracy of Equation (1)

R-Squared (R^2)	0.11
Standard Error	0.94
Standard Deviation	26.9

The coefficient of determination (R^2) is an indicator that ranges from 0 to 1 and which measures how closely the estimated values correspond to actual values. A trendline is most reliable when its R-squared value is at or near 1. The *Standard error,* i.e. *standard deviation dispersion,* of the target variable is a dimensionless measure of the model accuracy with values within the interval of [0,1] where the value of 0 corresponds to a 100% accurate prediction, and the *Standard deviation* is the overall accuracy of model prediction in the units of the target variable.

Figure 5.1 compares the process times generated by the VM Milling process with the estimated times derived through use of Equation (1).

Figure 5. 1: Stepwise Linear Regression-Estimated Vs Actual Process Times

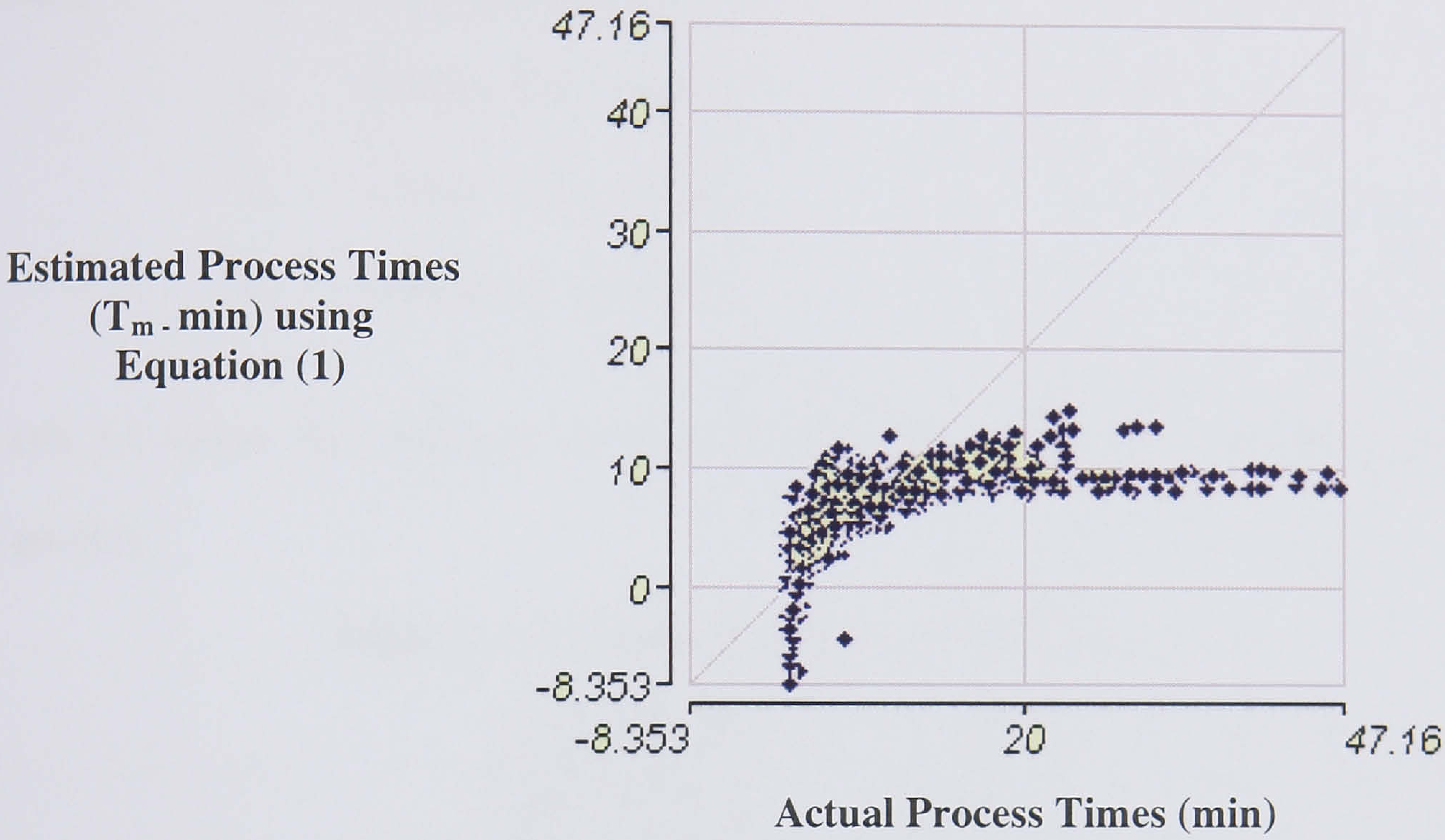


Figure 5.1 shows the Estimated vs. Actual process times for milling process data set and shows that the actual milling process times is not accurate with estimated values, i.e. the more accurate the estimated model, the closer the data points will be to the central diagonal. In addition, the appearance of curve in Figure 5.1 indicates the presence of non-linearity and hence predictions made using the SLR model significantly underestimated when actual process times are high or low i.e. for process time except in the range of 1 to 10 minutes.

5.1.1.3 Find Laws Algorithm

The following model for estimating T_m was developed using the Find Laws Algorithm, i.e.

$$T_m = 0.97 \times L_c V_f^{-1} \qquad \textbf{(2)}$$

Where: T_m = Cycle time (min)

L_c = Machining length (mm)

V_f = Feed rate (mm/min)

F_t = Feed per tooth (mm)

Table 5.3 shows the estimating accuracy of Equation (2) obtained using the Find Laws algorithm.

Table 5. 3: Estimating Accuracy of Equation (2)

R-Squared (R^2)	0.99
Standard Error	0.02
Standard Deviation	0.83

Figure 5.2 provides a comparison of actual process times versus estimated times obtained using the model derived from the Find Laws algorithm.

Figure 5. 2: Find Laws Algorithm-Estimated Vs Actual Process Times

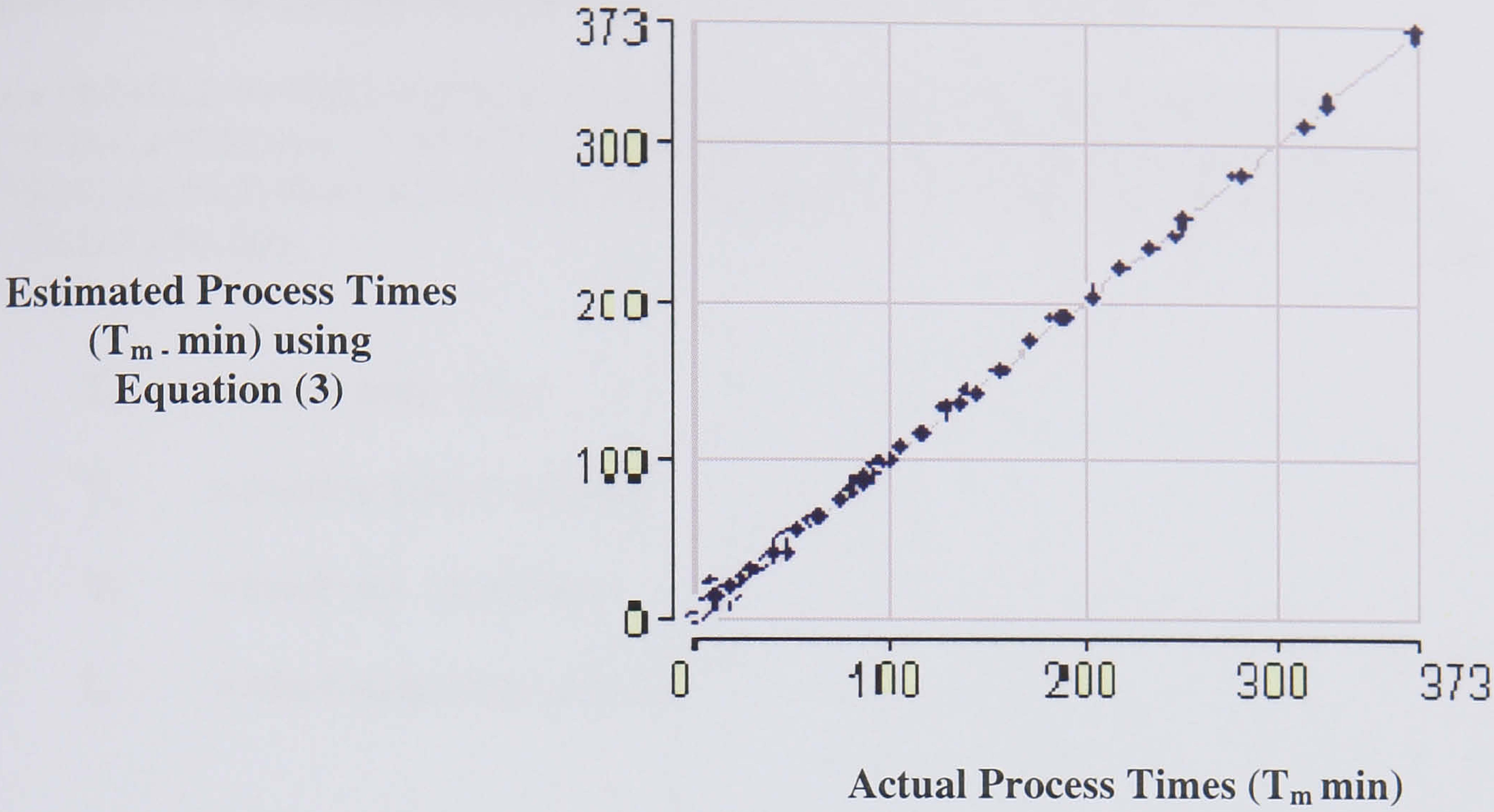


Figure 5.2 shows the Estimated vs. Actual process times, for the milling process data set, plotted using the FL algorithm and shows that estimated values have a high degree of accuracy. The resulting equation (Equation 2) indicates the existence of a strong non-linear relationship between the independent variables and the milling machining times.

5.1.1.4 PolyNet Predictor Algorithms

PolyNet predictor is an Artificial Neural Network based algorithm which when trained represents a hierarchical network that comprises of nodes and layers in the form of the hidden relationships found between the dependent and independent variables. It primarily produces output rules with 2nd or 3rd degree polynomials if the network is small. However, it does not display the results in a compact and readable form for large neural networks, which is one of the limitations of the artificial neural network approach. The network output from the PolyNet predictor for the milling process consists of 3 layers and 12 nodes.

Equation (3) is the polynomial expression for this three-layer, 12-node network.

Tm = (0.2+(2.2-Vc*0.02+Lc*0.2)*(-0.2+2.2-Vc*0.02+Lc*0.2)*(0.2+0.07*(2.2-Vc*0.02+Lc*0.2))))+(1.7-Vf*0.09)+Lc*0.2)*(0.7+(1.7-Vf*0.09+Lc*0.2))*(-0.07*(1.7-Vf*0.09)+Lc*0.2)+0.4*(2.2-Vc*0.02 +Lc*0.2))+(2.2-Vc*0.02+Lc*0.2)*(-0.2-0.3*(2.2-Vc*0.02+Lc*0.2)))) **(3)**

Where:

T_m = Cycle time (min)

V_c = Surface Speed (m/min)

V_f = Feed rate (mm/min)

L_c = Machining length (mm)

Table 5.4 shows the estimating accuracy of the network rule obtained using the PolyNet Predictor algorithm for the milling process.

Table 5. 4: Estimating Accuracy of Equation (3)

R-Squared	0.78
Standard Error	0.46
Standard Deviation	13.33

Figure 5.3 illustrates the graph of actual process times versus estimated process time derived through Equation (3).

Figure 5. 3: PolyNet Predictor Algorithm- Actual Vs Estimated Process Times

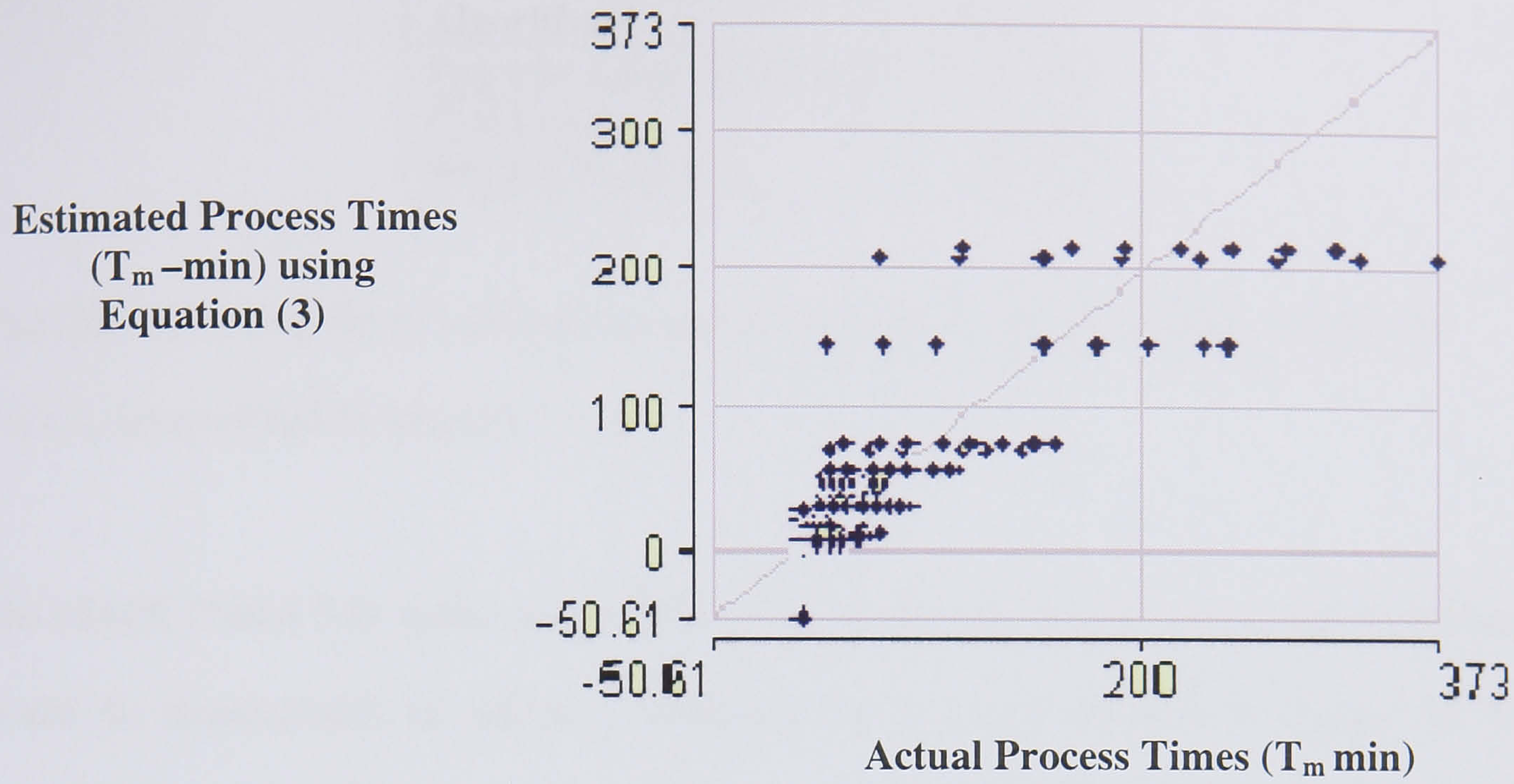


Figure 5.3 shows the Estimated vs. Actual process times for the milling process data set plotted using the PnP algorithm and shows that the estimated milling process times are inaccurate, i.e. the majority of the process time values are less than 100 mins and are densely located near the diagonal. The process times greater than 100 mins are dispersed away from the diagonal appears in the form of layers. Further analysis of milling results is discussed in Section 6.3.3.1.

Table 5.5 lists the Percentage Average Absolute Error (MAPE) obtained for the milling cost models through use of Equations (1), (2) and (3). MAPE is the average value of the absolute value of errors expressed in percentage terms.

It is calculated as $\dfrac{|\text{Actual - Estimated}|}{\text{Actual}} \times 100$

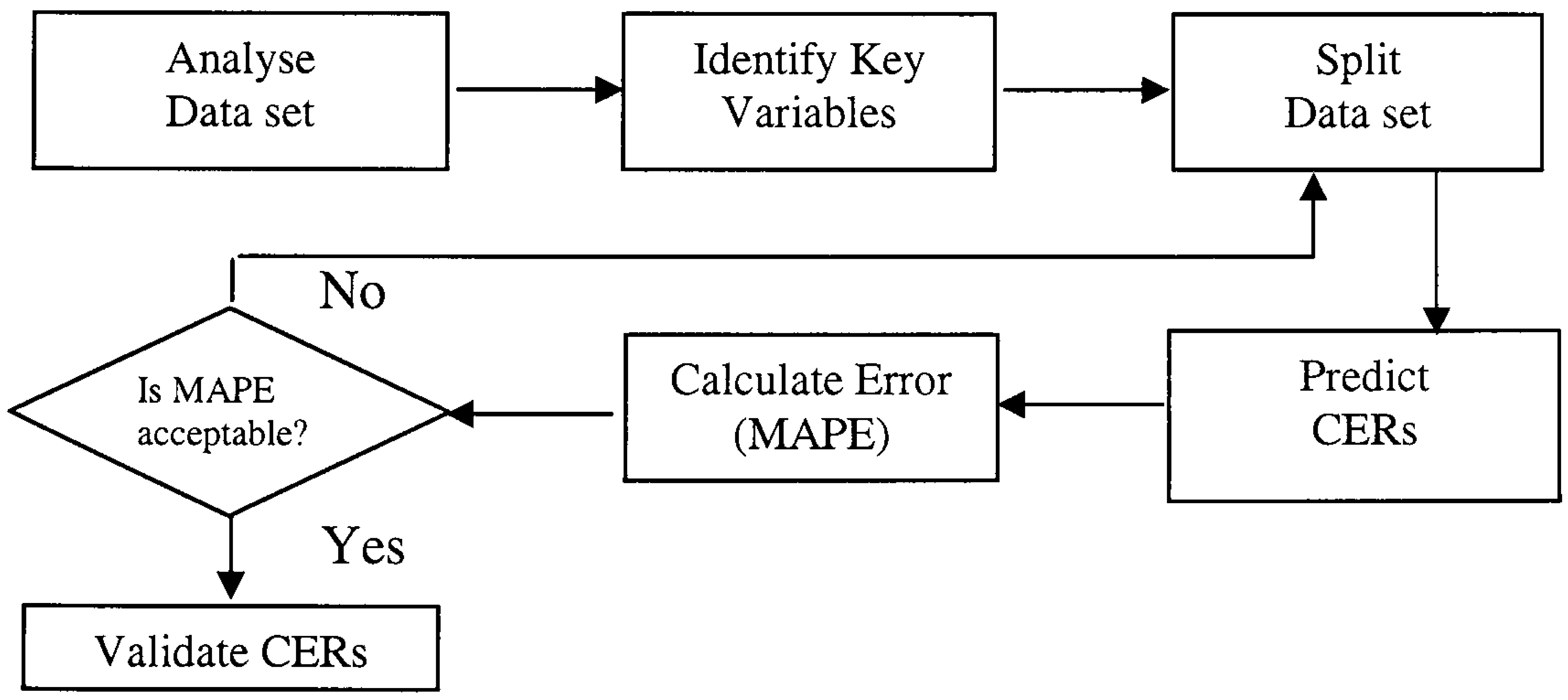
Table 5. 5: MAPE for Milling Models

Algorithms	MAPE
Stepwise Linear Regression	1065.53
Find Laws	31.88
PolyNet Predictor	921.73

The CER modelling for the milling data points excluded by the FD process, i.e. Section 5.1.1.1, are provided in Appendix 5.1A.

The MAPE (Table 5.5) values of the CER's for the milling process were high and hence would be unacceptable in practice. Although, the accuracy obtained by using the FL approach is considerably better when compared to the accuracies of SLR and PnP. It is still not sufficient to be acceptable in practice. The dataset should, therefore, be analysed in further detail in order to identify the root cause of this estimating inaccuracy. A method, therefore, is proposed (see Figure 5.4), for resolving this issue without recourse to the use of the FD algorithm.

Figure 5. 4: Method to Improve CERs MAPE



The first step of this method is to identify the key variables contributing to the resulting CER. From Equations (1), (2) and (3) it is apparent that the accuracy of the models developed using FL is more better than the other two methods, and in that Equation (2) V_f and Lc are the variables that have the greatest influence on the estimating accuracy of the resulting CERs. Analysis of the milling data set revealed that of these variables V_f had a wider range of values and higher levels of variation through out its entire data set.

The next step involved grouping the data set based on the most significant variable V_f. This involved, the whole milling data set, (1679 data points) being grouped into data sub-sets based on the percentage values of V_f shown in Table 5.6.

Table 5. 6: Summary of Distribution of Data Points based on % V_f in Milling Data set

Cumulative % of V_f	No. Data Points	V_f Min	V_f Max	V_f Range
5	678	1.32	17.19	15.78
10	477	22.00	34.38	12.38
15	244	36.67	53.48	16.81
20	145	57.30	68.76	11.46
25	70	76.40	91.68	15.28
50	49	114.60	171.09	56.49
Rest (50)	15	183.36	366.72	213.36

Table 5.6 shows the distribution of V_f in the whole data set where a Pareto distribution is evidenced by the distribution of V_f where in 5% of V_f appear to have more influence on estimating accuracies when compared to 95% of V_f. Hence, two new datasets (A_a and A_b) with 5% and 95% of V_f were created as shown in Table 5.7.

Table 5. 7: Two Data Sub-Sets using V_f

Data set	%of Vf	No. of Data points
A_a	5	678
A_b	95	1001

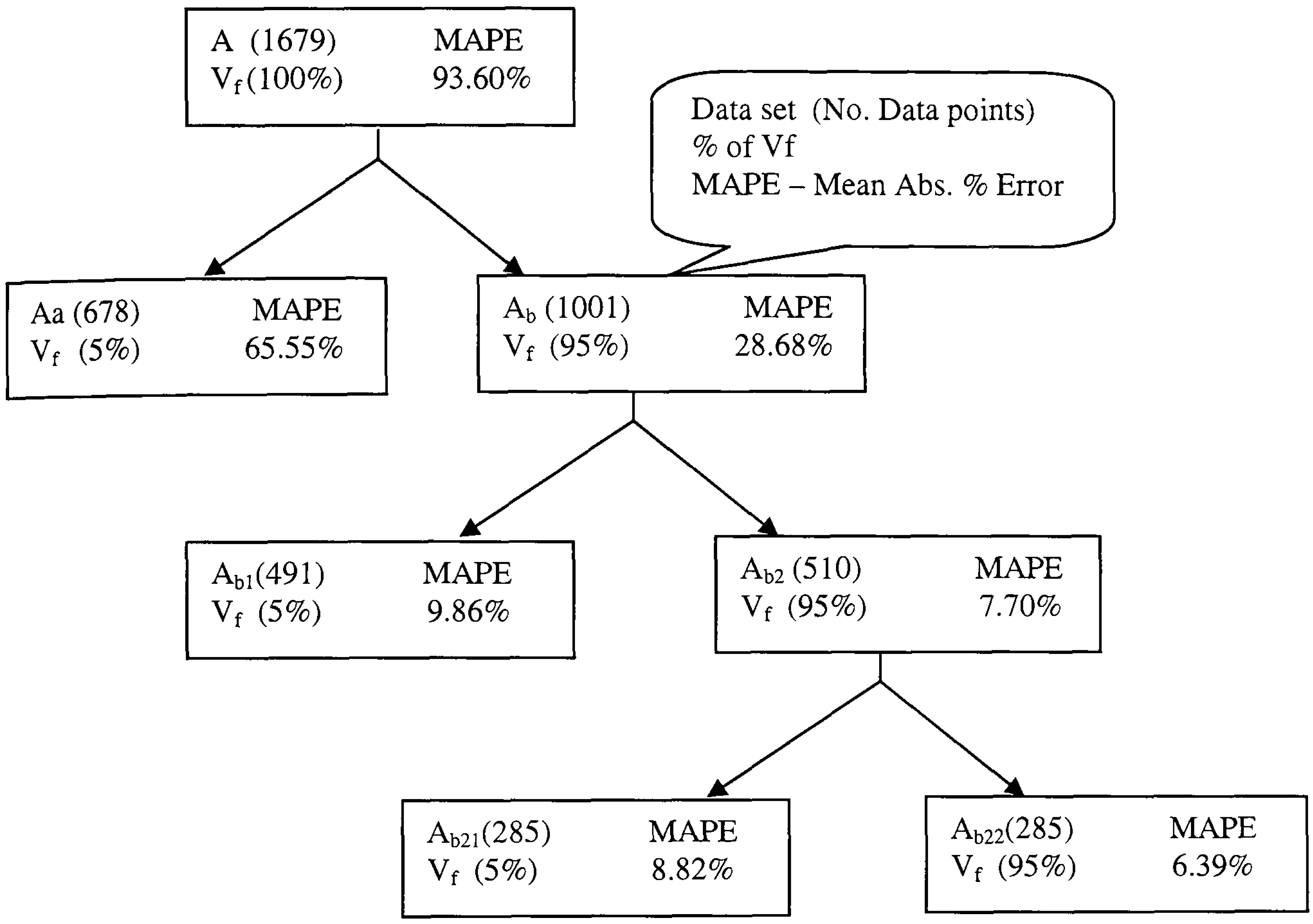
Experiments were carried using these data sets using the Find Laws algorithm to establish the CERs. The results obtained from this experimentation are listed in Table 5.8.

Table 5. 8: Result of Milling Data sets

Data set	% V_f	CER	MAPE
A_a	5	$T_m = L_c(0.58 - 2.11\times F_t)\ (V_f^2 D_c F_t)^{-1}$	66.55%
A_b	95	$T_m = (0.52 + 0.02\times L_c V_f)(V_f - 20.86 + 3.07\times D)^{-1}$	28.68%

Results obtained from experiments on the two data sub-sets, i.e. $\mathbf{A_a}$ and $\mathbf{A_b}$, revealed that the MAPE values had been greatly reduced but not sufficiently for practical estimating purposes. Hence, in order to achieve a lower value of MAPE, the Pareto Rule was applied to data set $\mathbf{A_b}$ by splitting this data set in to two further data sub-sets namely $\mathbf{A_{b1}}$ and $\mathbf{A_{b2}}$. This process of data sub-set grouping was repeated until sufficient estimating accuracy as indicated in the values of MAPE was achieved, see Figure 5.5.

Figure 5. 5: Milling Split Dataset with MAPE



5.1.2 Automated Paint Spraying

Experiments were carried out using the data generated from the automated paint spraying process models described in Section 4.2.3. The results of these experiments are listed in Tables 5.9 to 5.13. Identification of the CERs within the data range identified through the use of Find Dependencies algorithm was then carried out to compare the use of the predictive data mining algorithms Stepwise Linear Regression, Find Laws and PolyNet Predictor.

5.1.2.1 Find Dependencies

Experiments were carried out using the dataset obtained from the Automated Paint Spraying virtual process models described in Section 4.2.3. Initially, the dataset was explored using the Find Dependencies algorithm and the results of these experiments are shown in Table 5.9.

Table 5. 9: Results from Find Dependencies Algorithms

FD Outputs	**Results**
Size (No. of Data Points)	1188
IPV	D_{wp} (Distance between paint gun and work piece), P_{fr} (Paint fluid ratio), G_s (paint gun speed), G_{rg} (Paint gun range), P_{tl} (path length), P_{ns} (paint sprayed), P_{nw} (Paint wasted)
DPV	Tm
MIV	G_s, P_{tl}
Size (No. of Data Points)	1148
P-value	0

5.1.2.2 Stepwise Linear Regression

Equation (4) for Automated Paint Spraying was developed using Stepwise Linear Regression;

$$T_p = (7.9 - 0.002D_{wp} - 0.006P_{fr} - 0.007G_s - 0.0007G_{rg} + 0.001P_{tl} + 0.07P_{ns} - 0.005P_{nw}) \quad \textbf{(4)}$$

Where:

T_p = Cycle time (min)

Dwp = Distance between paint gun and work piece (mm)

P_{fr} = Paint fluid flow rate (cc/min)

G_s = Paint gun speed (mm/sec)

G_{rg} = Paint gun range (mm)

P_{tl} = Path length (mm)

P_{ns} = Paint sprayed (cc)

P_{nw} = Paint wasted (cc).

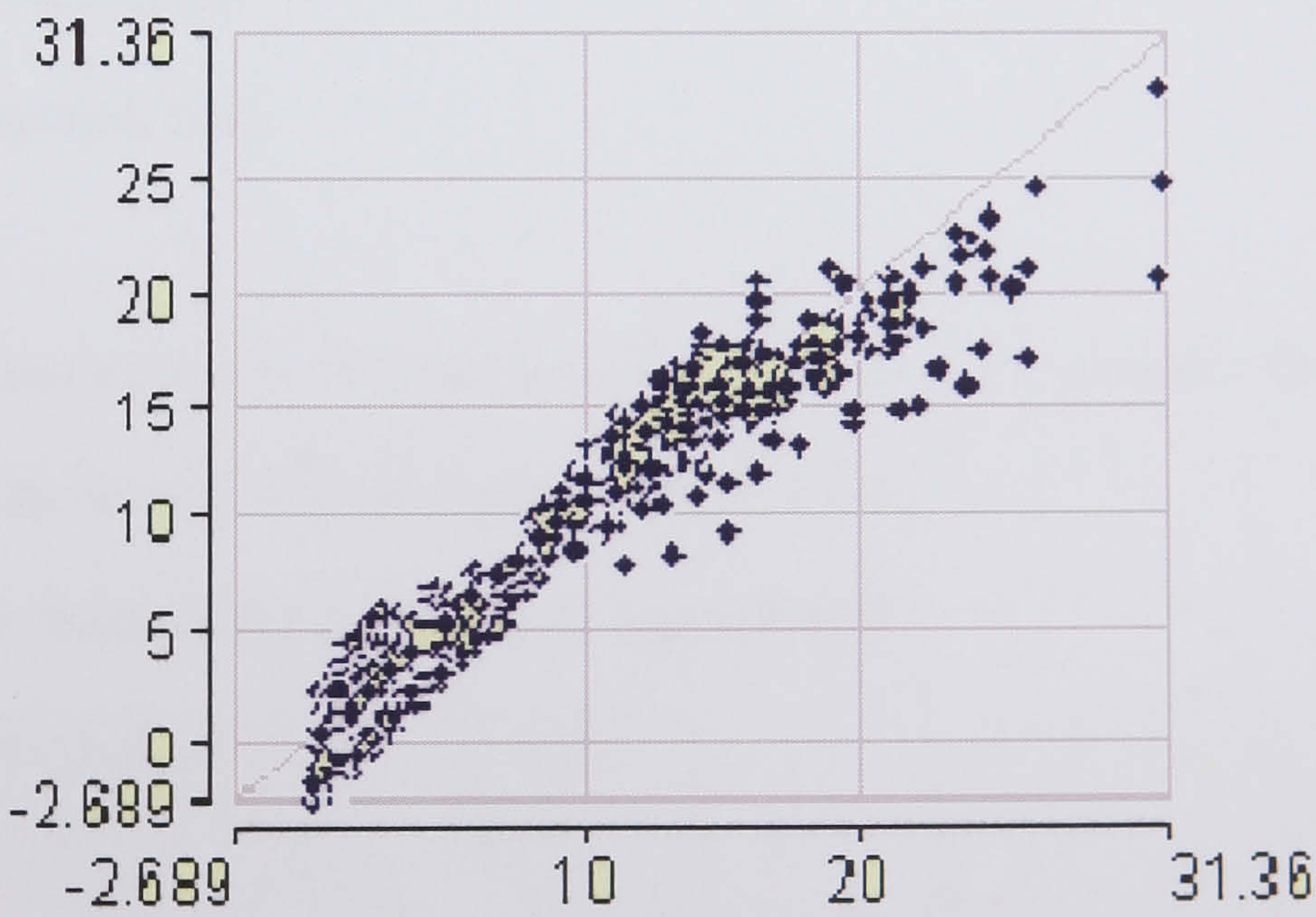
Table 5.10 provides values for the measures of accuracy and significance of the model developed and Figure 5.6 compares for the automated paint spraying actual versus the estimated process times generated using Equation (4).

Table 5. 10: Estimating Accuracy of Equation (4)

R-Squared (R^2)	0.92
Standard Error	0.27
Standard Deviation	1.47

Figure 5. 6: Stepwise Linear Regression- Actual Vs Estimated Cycle Times

Estimated Process Times (T_p-min) using Equation (4)



Actual Process Times (T_p-min)

Figure 5.6 shows the Estimated vs. Actual process times for the Automated Paint Spray process data set and shows that the estimated values are inaccurate, i.e. the process time above 10 minutes are underestimated and are lying below the diagonal. This dispersion indicates towards the existence of non-linear relationship between paint process time and its independent process variables.

5.1.2.3 Find Laws

The following model for Automated Paint Spraying was developed using the Find Laws Algorithm,

$$T_p = (P_{tl}(0.65 + 0.002G_s) + 68.17P_{ns}))(G_s + 0.92P_{fr})^{-1} \quad \textbf{(5)}$$

Where: C_{tm} = Process cycle time (min)

P_{fr} = Paint fluid flow rate (cc/min)

G_s = Paint gun speed (mm/sec)

P_{tl} = Path length (mm)

P_{ns} = Paint sprayed (cc).

Table 5.11 shows the estimating accuracy of Equation (5) and Figure 5.9 compares for automated spray paint process the actual versus estimated process times.

Table 5. 11: Estimating Accuracy of Equation (5)

R-Squared (R^2)	0.98
Standard Error	0.10
Standard Deviation	0.55

Figure 5. 7: Find Laws Algorithm- Actual Vs Estimated Process Times

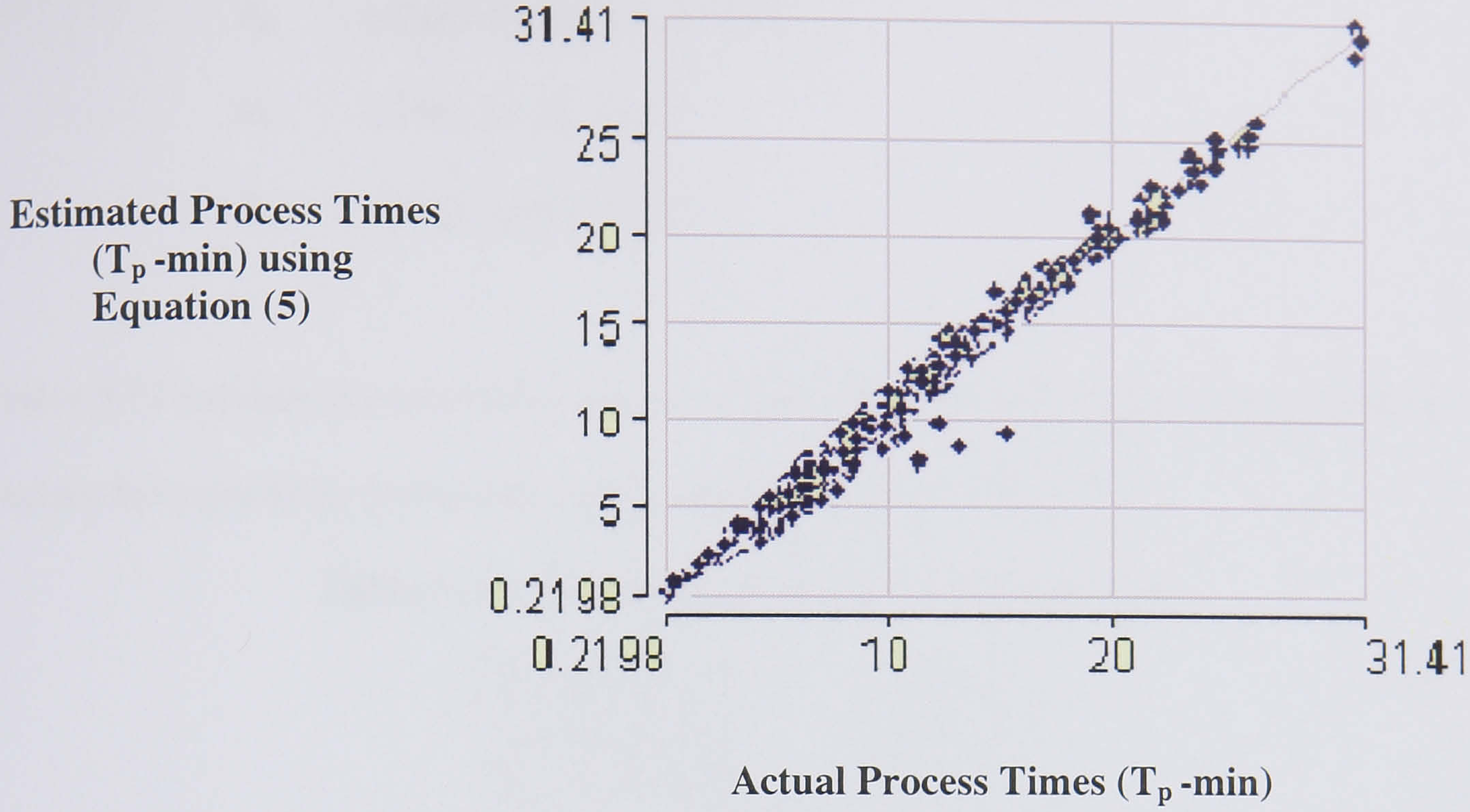


Figure 5.7 shows the Estimated vs. Actual process times for the automated paint spray process data set plotted using the FL algorithm and shows that the estimated values have high degree of accuracy. The resulting equation (Equation 5) indicates the existence of a strong non-linear relationship between the independent variables and the automated paint spray times.

5.1.2.4 PolyNet Predictor

Equation (6) illustrates the polynomial expression developed for Automated Paint Spraying from a two-layer 7-node network using the PolyNet predictor algorithm.

$$C_{tm} = (0.072 +(-P_{fr}*0.009 +P_{ns}*(0.38- P_{ns}*0.001)*(-0.05+(- P_{fr}*0.009 + P_{ns}*0.38) *(0.05-0.0035*(P_{ns} *0.38))) +(P_{tl}*0.004)*(0.95+(P_{tl}*0.004)*(-0.03 +0.001 * (P_{tl} *0.004)+0.002*(P_{fr}*0.009) + P_{ns} *(0.38+ P_{ns} *0.001- P_{fr}*0.0004)))-(P_{fr}*0.009) + P_{ns} *(0.38- P_{ns}*0.001- P_{fr}*-0.0004))*(-0.02+0.005*(- P_{fr}*0.009 + P_{ns} *(0.38- P_{ns} *0.001 + P_{fr} *0.0004))))) \quad (6)$$

Where: C_{tm} = Cycle time (min)

P_{fr} = Paint fluid flow rate (cc/min)

P_{tl} = Path length (mm)

P_{ns} = Paint sprayed (cc).

Table 5.12 indicates the estimating accuracy of Equation (6) and Figure 5.8 compares for automated paint spray process the actual versus estimated process times.

Table 5. 12: Estimating Accuracy of Equation (6)

R-Squared (R^2)	0.96
Standard Error	0.17
Standard Deviation	0.95

Figure 5. 8: PolyNet Predictor- Actual Vs Estimated Process Times

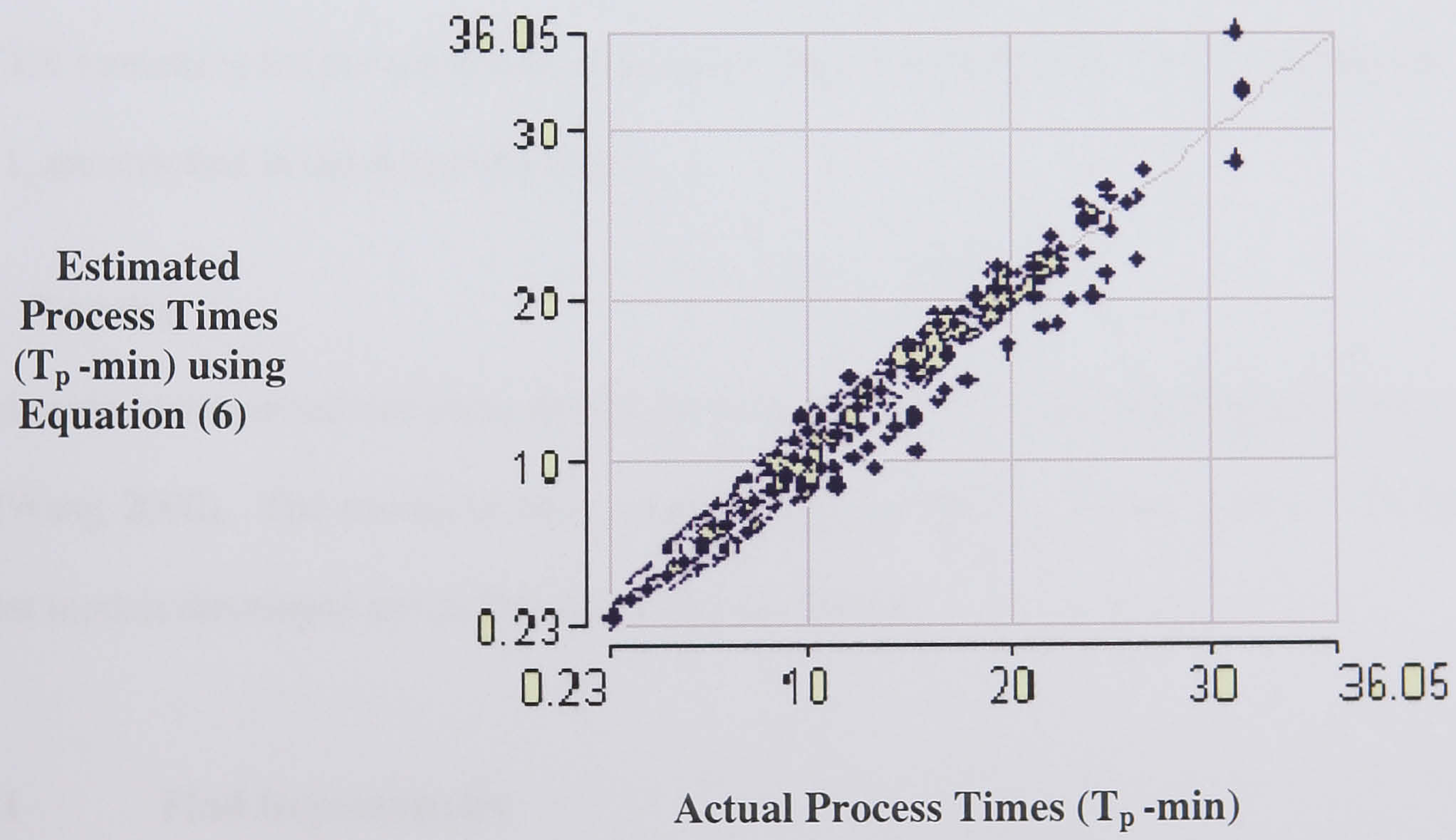


Figure 5.8 shows the Estimated vs. Actual process times for the automated paint spray process data set plotted using the PnP algorithm. This figure shows that the data points are distributed closely along the diagonal, which indicates the ability of PnP to deal with non-linearity in the data points. However, the appearance of distance between data points and diagonal is due to the existence of standard error (0.17).

Table 5.13 lists the MAPE obtained for the automated spray painting cost models through use of Equations (4), (5) and (6).

Table 5. 13: MAPE for Automated Paint Spraying

Algorithms	MAPE
Stepwise Linear Regression	30.2
Find Laws	5.1
PolyNet Predictor	7.7

The CER modelling for the automated paint spray data points excluded by FD i.e. Section 5.1.2.1, are attached in the Appendix 5.2A.

5.1.3 Turning

Experiments were carried out using the data generated from the rough turning experimental data (Wang 2000). The results of these experiments are listed in Tables 5.14 to 5.18 and the cost models developed are in Equations (7), (8) and (9).

5.1.3.1 Find Dependencies

Initially, the collected dataset was explored using the Find Dependencies Algorithm with the results listed in Table 5.14.

Table 5. 14: Outputs from Find Dependencies Algorithms

FD Outputs	Results
Size (Actual No. Data Points)	750
IPV	r_v, ps, W, d_m, R_{sg}, r_i, l_r, r_e, n, t_{sa}, n_t, t_{sb}, B_s, t_{ln}, n_o, t_{pt}
DPV	T_m
MIV	n_t, B_s, t_{ln}, t_{pt}
Size (No. Data Points)	440
P-value	0

The data selected by the Find Dependencies algorithm was then analysed using the predictive data mining algorithms of Stepwise Linear Regression, Find Laws and PolyNet Predictor.

5.1.3.2 Stepwise Linear Regression

Equation (7) for Turning was developed using Stepwise Linear Regression:

$$\begin{aligned} T_t = (&51.59 + 18.64r_v + 0.016t_{sa} + 7.18n_t \\ &+ 0.07t_{sb} - 1.60B_s + 1.23t_{ln} + 4.18n_o + 4.09t_{pt}) \end{aligned} \qquad \textbf{(7)}$$

Where:

T_t = Cycle time (min)

r_v = Proportion of initial volume

t_{sa} = Basic set-up time for machine (min)

n_t = Number of tools

t_{sb} = Set-up time per tool (min)

B_s = Batch size

t_{ln} = Loading and unloading time (min)

n_o = Number of operations

t_{pt} = Tool positioning time per operation (min).

Table 5.15 provides the estimating accuracy of Equation (7) and Figure 5.7 compares for the Turning operation the Actual Versus Estimated Process times derived from Equation (7).

Table 5. 15: Estimating Accuracy of Equation (7)

R-Squared (R^2)	0.56
Standard Error	0.65
Standard Deviation	46.9

Figure 5. 9: Stepwise Linear Regression –Actual Vs Estimated Process Times

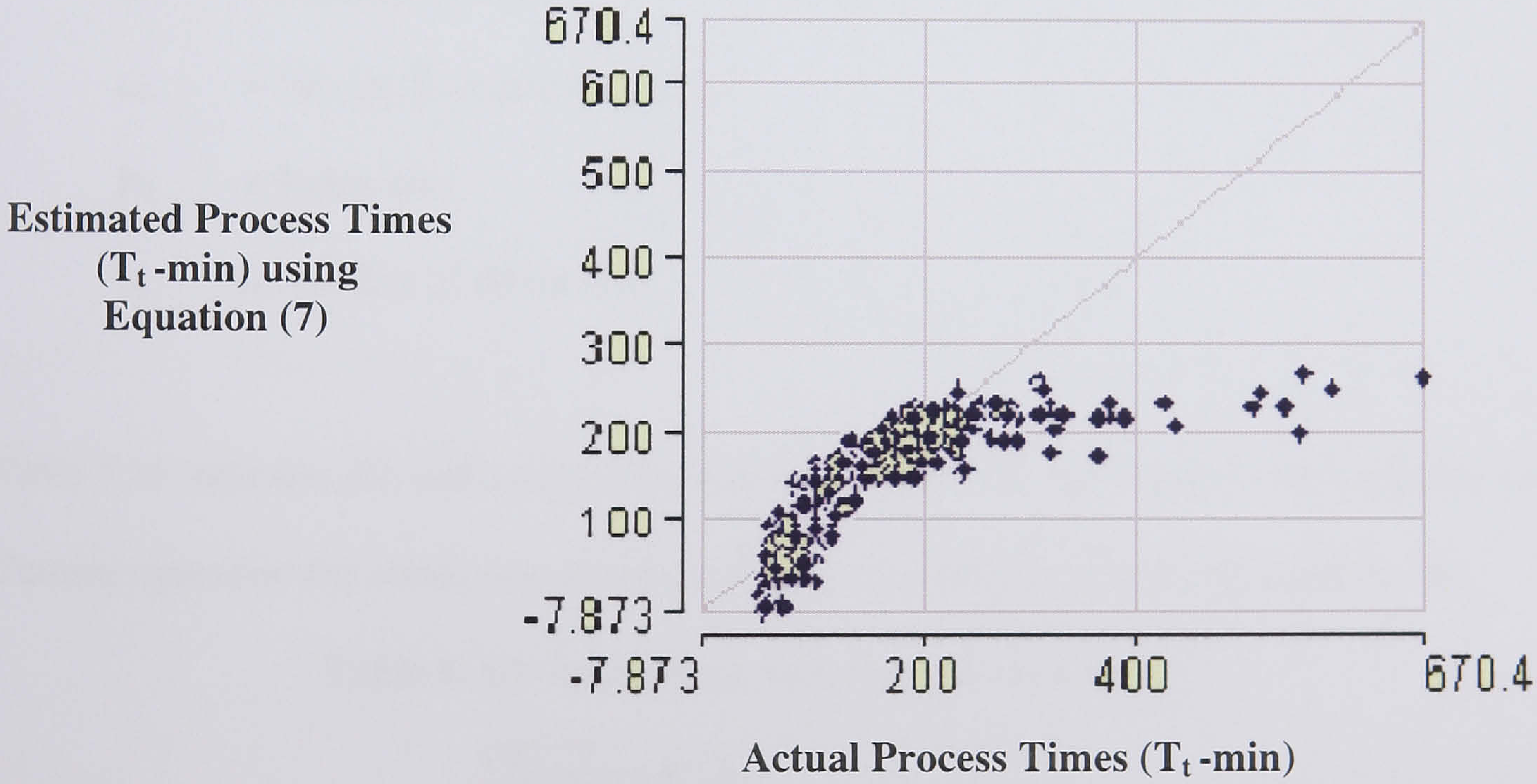


Figure 5.9 shows the Estimated vs. Actual process times for the turning process data set and shows that the estimated turning process times are inaccurate. The appearance of curve in

this figure indicates the non-linearity, which can be predicted more accurately using non-linear algorithms such FL and PnP. Hence the predictions made using the SLR model significantly underestimate when actual process times are high or low i.e. for actual process times above 200 minutes.

5.1.3.3 Find Laws

Equation (8) for Turning was developed using the Find Laws Algorithm, i.e.:

$$T_t = (41.86B_s + 1.13t_{sa} + 1.07n_t t_{sb} + 5.1n_o Bs + 63.72n_o)(B_s+0.7)^{-1} \quad \textbf{(8)}$$

Where:

T_t = Cycle time (min)

t_{sa} = Basic set-up time for machine (min)

n_t = Number of tools

t_{sb} = Set-up time per tool (min)

B_s = Batch size

n_o = Number of operations.

Table 5.16 indicates the estimating accuracy of Equation (8) and Figure 5.10 compares for Turning operation the actual versus estimated process times derived using Equation (8).

Table 5. 16: Estimating Accuracy of Equation (8)

R-Squared (R^2)	0.97
Standard Error	0.16
Standard Deviation	11.39

Figure 5. 10: Find Laws- Actual Vs Estimated Process Times

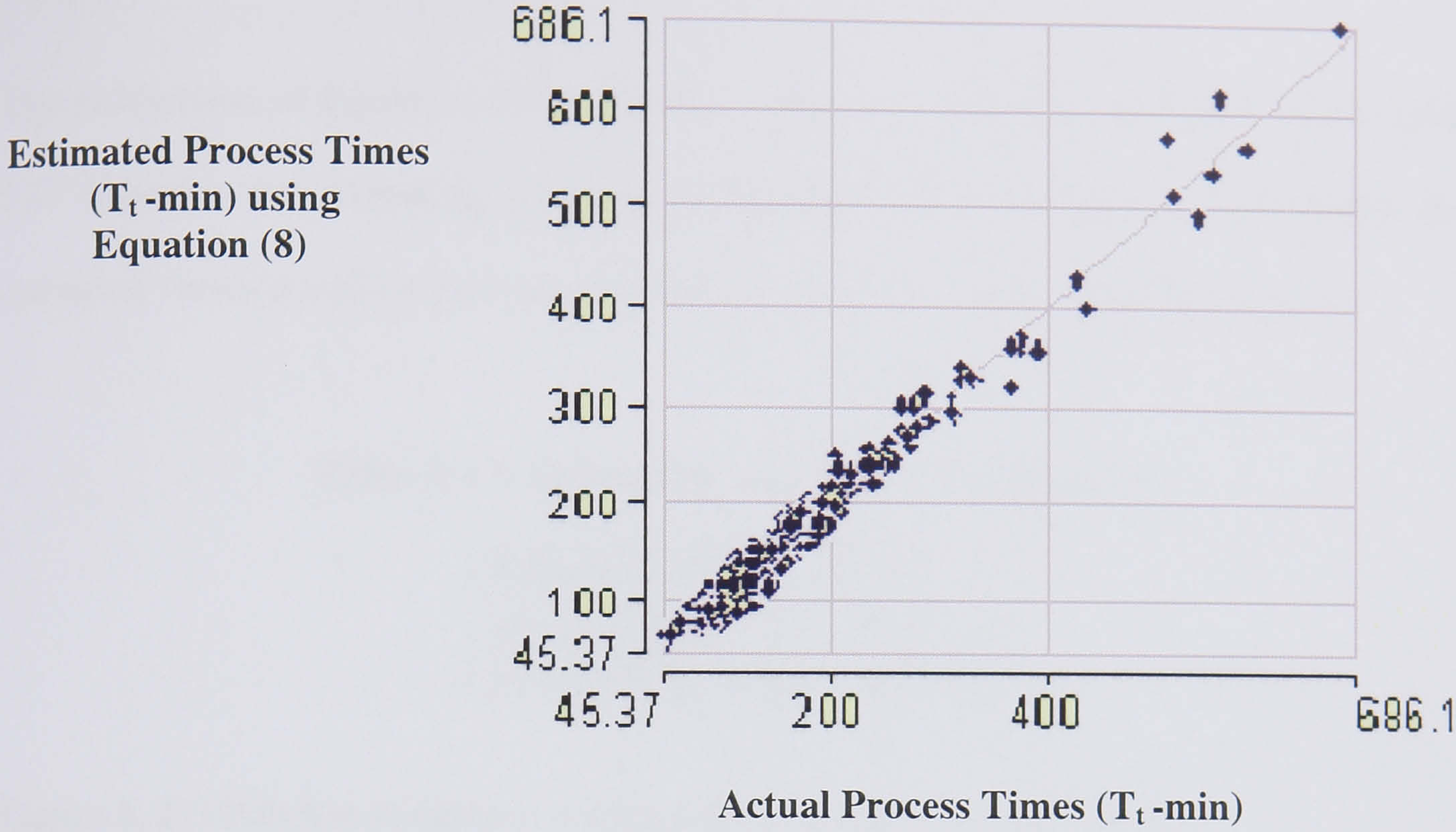


Figure 5.10 shows the Estimated vs. Actual process times for the turning process data set plotted using the FL algorithm and shows that the estimated values have high degree of accuracy. The resulting equation (Equation 8) indicates the existence of a strong non-linear relationship between the independent variables and the turning process times.

5.1.3.4 PolyNet Predictor

The following model for Turning was developed using the PolyNet Predictor Algorithm,

$$T_t = 235.38 + t_{sb}(0.53) + B_s(-8.80 + B_s(0.15 - 0.0008B_s + 0.0001t_{sb}) + t_{sb}(-0.02))) \quad \textbf{(9)}$$

Where:

T_t = Cycle time (min)

t_{sb} = Set-up time per tool (min)

B_s = Batch size.

The architecture of Equation (9) consists of 1 network layers and 3 network nodes. Table 5.17 indicates the estimating accuracy of Equation (9) and Figure 5.11 compares the estimated versus actual process times of turning operation derived from Equation (9).

Table 5. 17: Estimating Accuracy of Equation (9)

R-Squared (R^2)	0.77
Standard Error	0.47
Standard Deviation	33.98

Figure 5. 11: PolyNet Predictor - Estimated Vs Actual Process Times

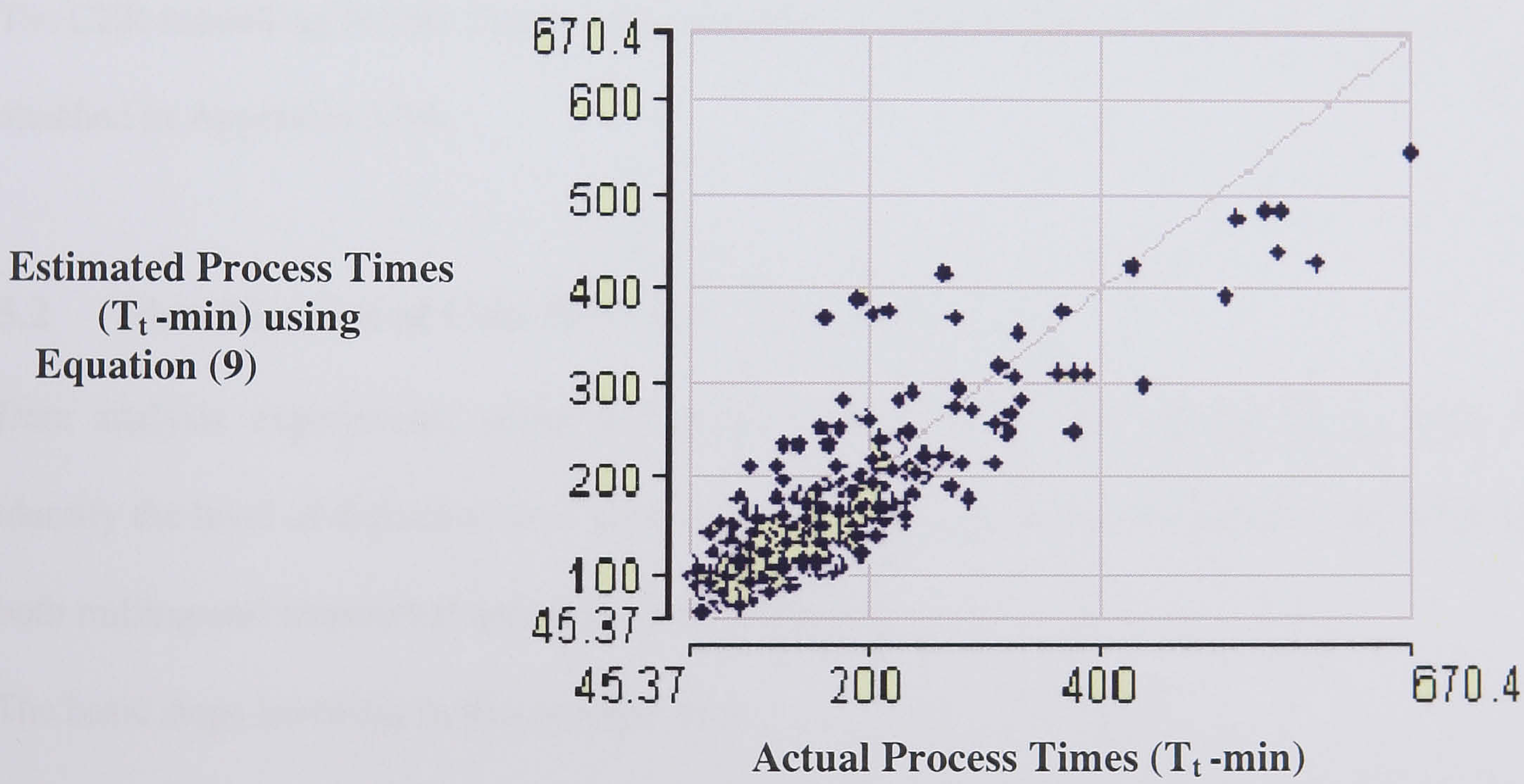


Figure 5.11 shows the Estimated vs. Actual process times for the turning process data set plotted using the PnP algorithm and shows that the actual turning process times are

scattered along the diagonal. This Figure indicates that the relationship is non-linear between turning process times and independent process variables. However, due to high value of standard error (i.e., 0.47) the estimated process times values are dispersed along the diagonal. Further analysis of turning results is discussed in Section 6.3.3.3.

Table 5.18 lists the MAPE obtained for the turning cost models through use of Equations (7), (8) and (9).

Table 5. 18: MAPE for Turning Process Models

Algorithms	MAPE
Stepwise Linear Regression	25.03
Find Laws	11.94
PolyNet Predictor	25.36

The CER modelling for the Turning data points excluded by FD i.e. Section 5.1.3.1, are attached in Appendix 5.3A.

5.2 Identification of Cost Drivers

Data analysis experiments using data mining algorithms were carried out in order to identify the level of dependence of dependent variables on individual predictor variables for both milling and automated spray-painting processes.

The basic steps involved in this process were:

(i) Select the predictor variable, and use the FD algorithm to identify the most influencing independent process variables on the selected predictor variable. These experiments provide outputs with a list of variables (i.e. cost drivers) and the data subset shown in Row 1, Table 5.19 and 5.20.

(ii) Repeat Step (i) for the remaining independent process variables to find their level of dependence on predictor variables, (i.e. Row 2, Table 5.19 and 5.20).

(iii) Repeat steps (i) and (ii) until all cost driver levels of dependence on predictor variable have been identified.

The results of these experiments in which the predictor variable is 'cycle times' are provided in Tables 5.19 and 5.20. The Find Dependencies results provide "the most influencing variables", (i.e. those variables found to be the most highly related to the dependent variables) and "number of points obeying the dependence", i.e. these are the number of data points in a particular dataset that follow the dependence discovered by FD algorithm.

Table 5. 19: Milling Cost drivers

Experiment Number	Cost Drivers	Data points obeying the dependence	Data points not obeying the dependence
1	Vf, Lc, Ft	1433	246
2	D, n	1223	461
3	Dc, Vc	865	819

Table 5. 20 : Automated Paint Spray Cost drivers

Experiment Number	Cost Drivers	Data points obeying the dependence	Data points not obeying the dependence
1	Gs, Ptl	1140	40
2	Pfr, Pns	1100	80
3	Grg, Pnw	970	210

5.3 Effect of Number of Variables and Data Points

This section provides the results of the experiments undertaken, (Section 4.5), to identify the effect of "Number of Variables" and "Number of Data Points" on model estimating accuracy for each of the manufacturing processes examined.

5.3.1 Model Accuracy vs Number of Data Points

The maximum "Number of Data Points" used to explore the effect of this variable on model estimating accuracy was 1200 for the Automated Spray Paint and 750 for the Turning process. Figures 5.12 to 5.14 provide the results of these experiments for the Automated Spray Paint process and Figures 5.15 to 5.18 for the Turning process.

Figure 5. 12: Average Accuracy Vs Number of Data Points

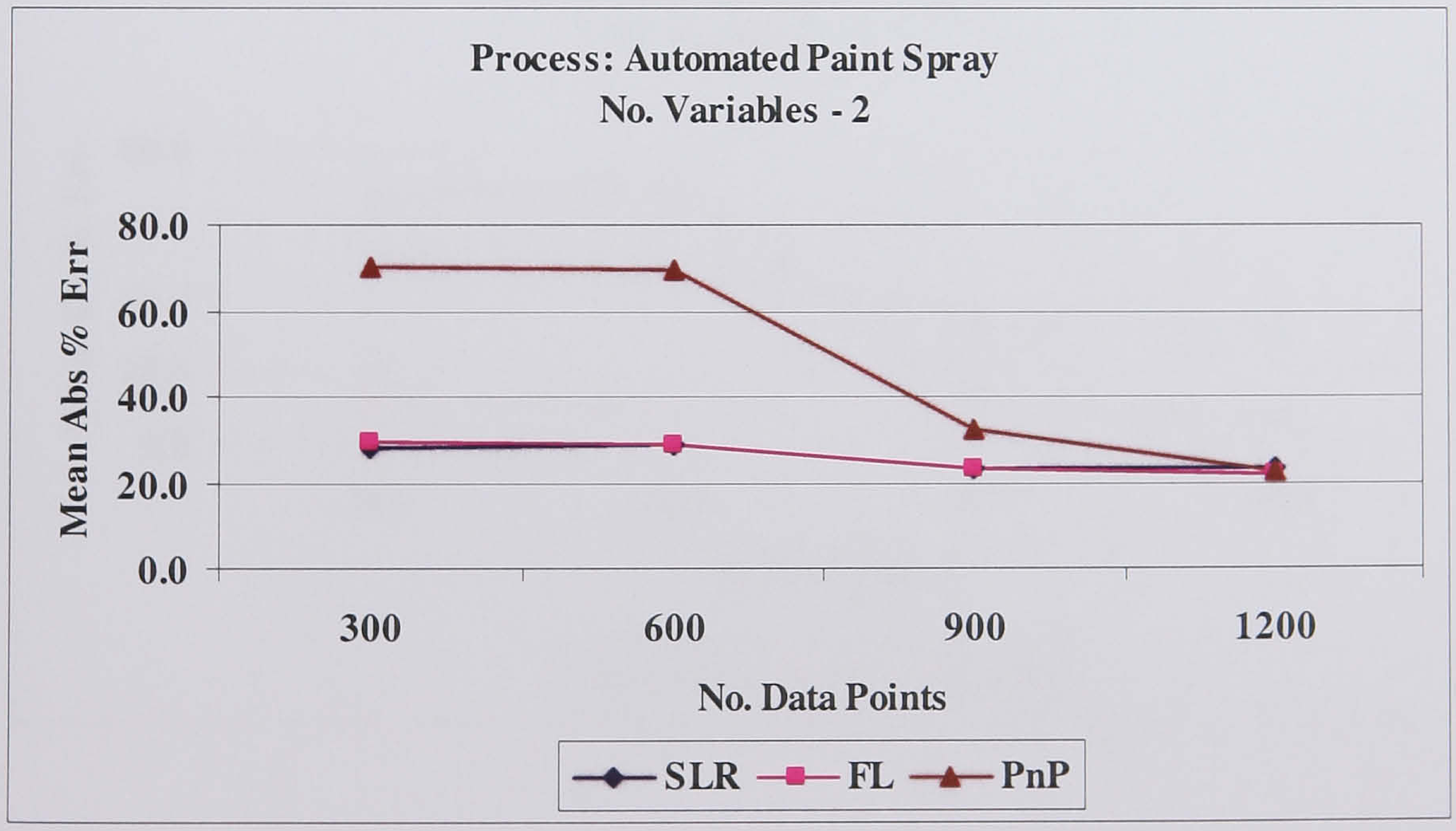


Figure 5. 13: Average Accuracy Vs Number of Data Points

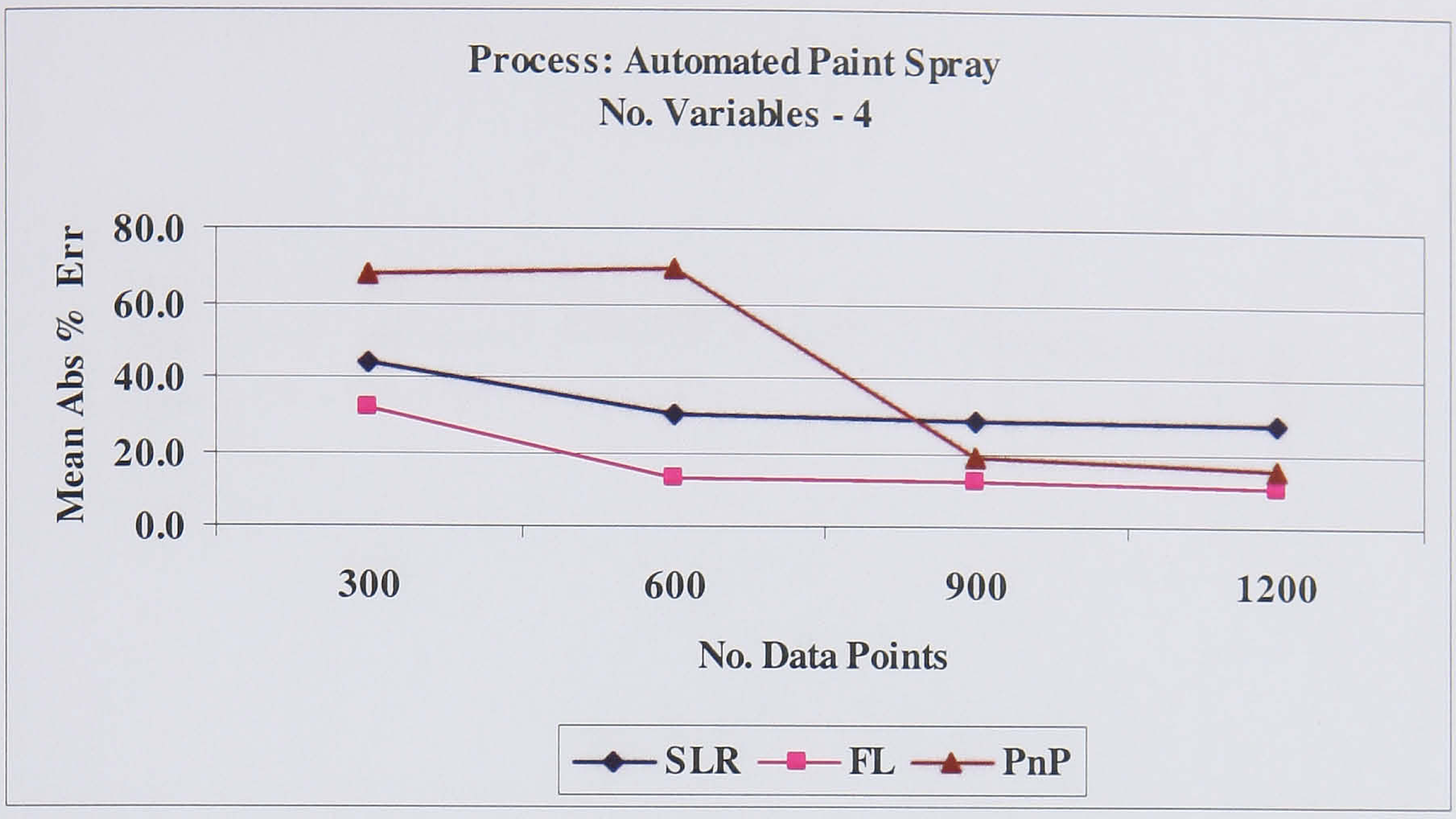


Figure 5. 14: Average Accuracy Vs Number of Data Points

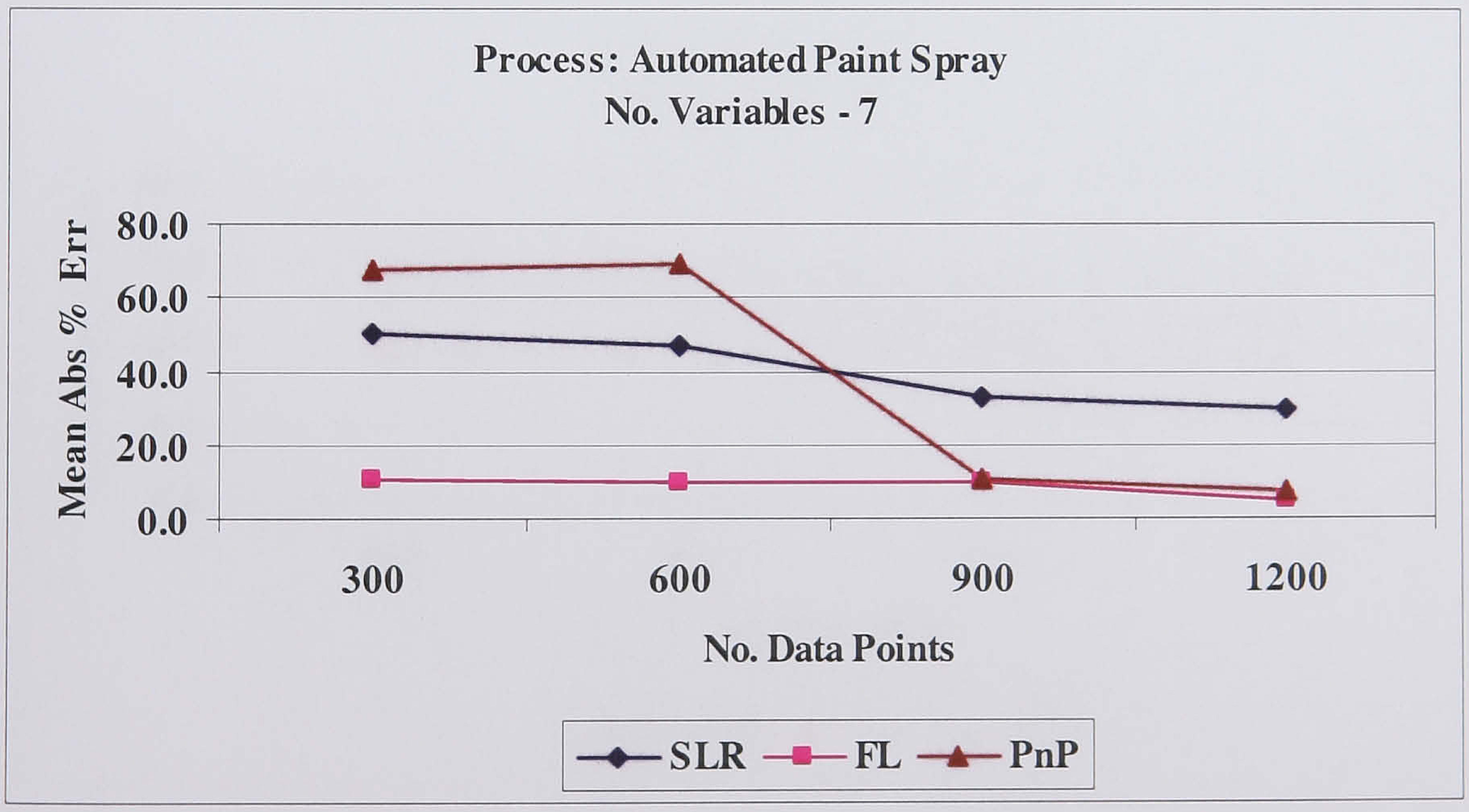


Figure 5. 15: Average Accuracy Vs Number of Data Points

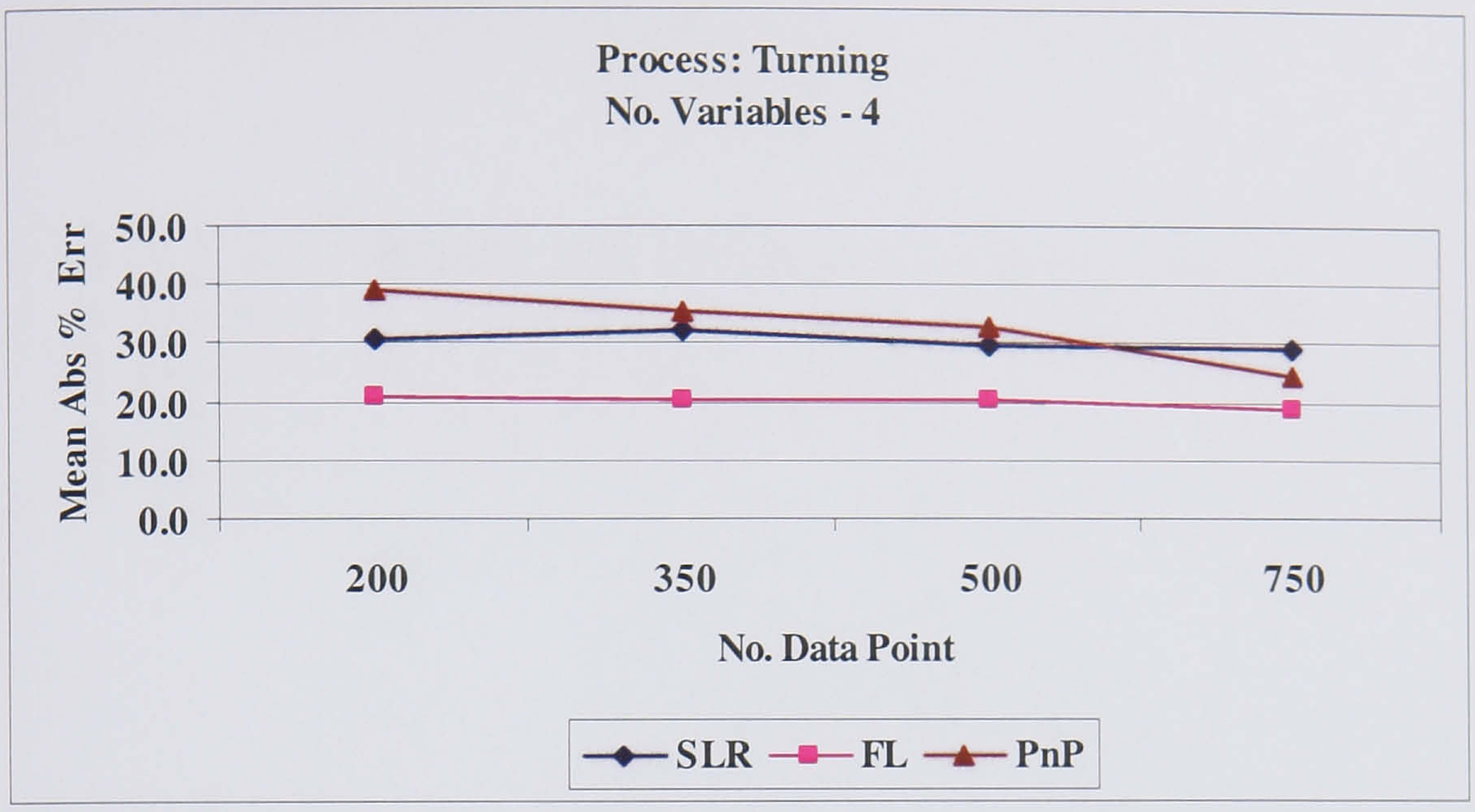


Figure 5. 16: Average Accuracy Vs Number of Data Points

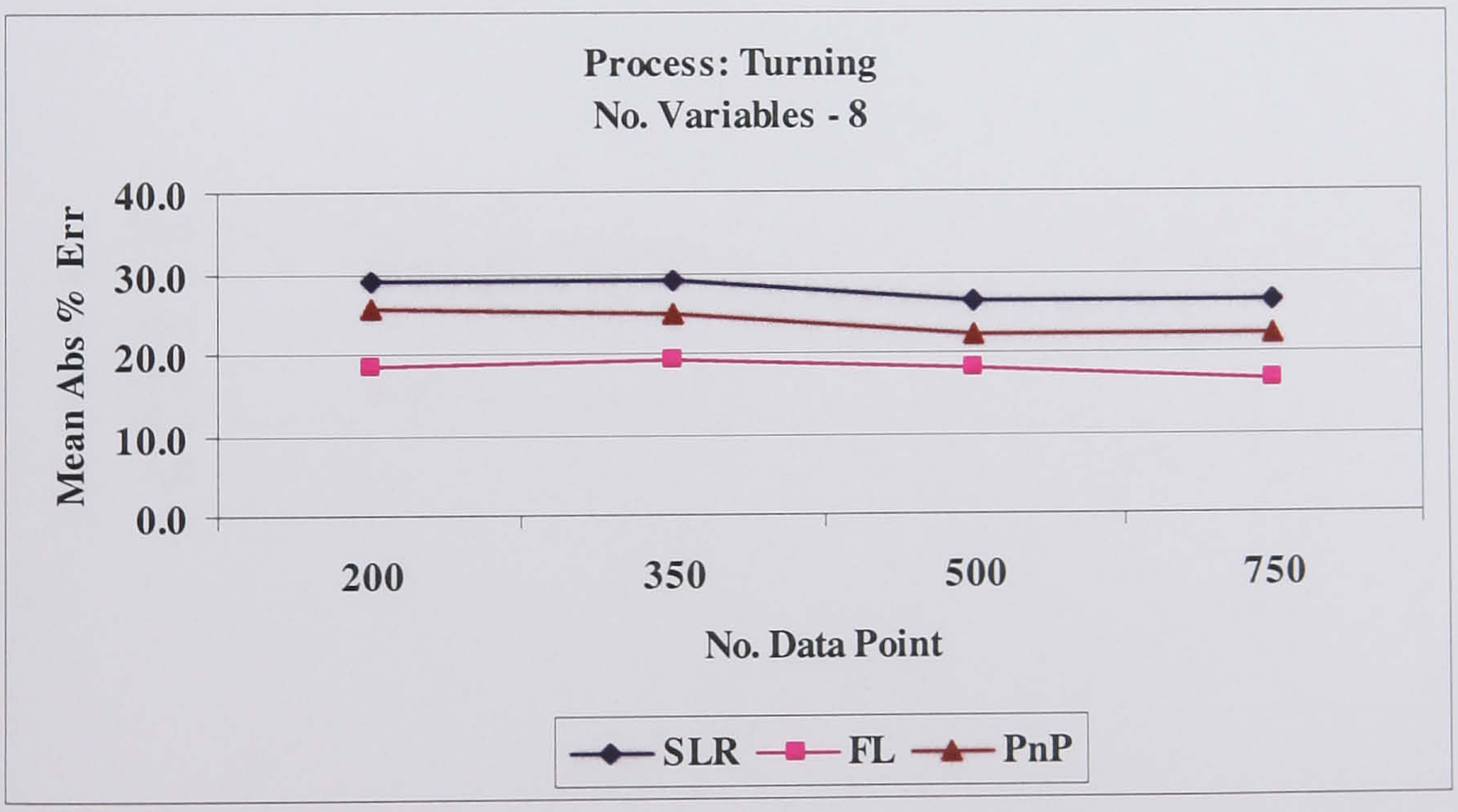


Figure 5. 17: Average Accuracy Vs Number of Data Points

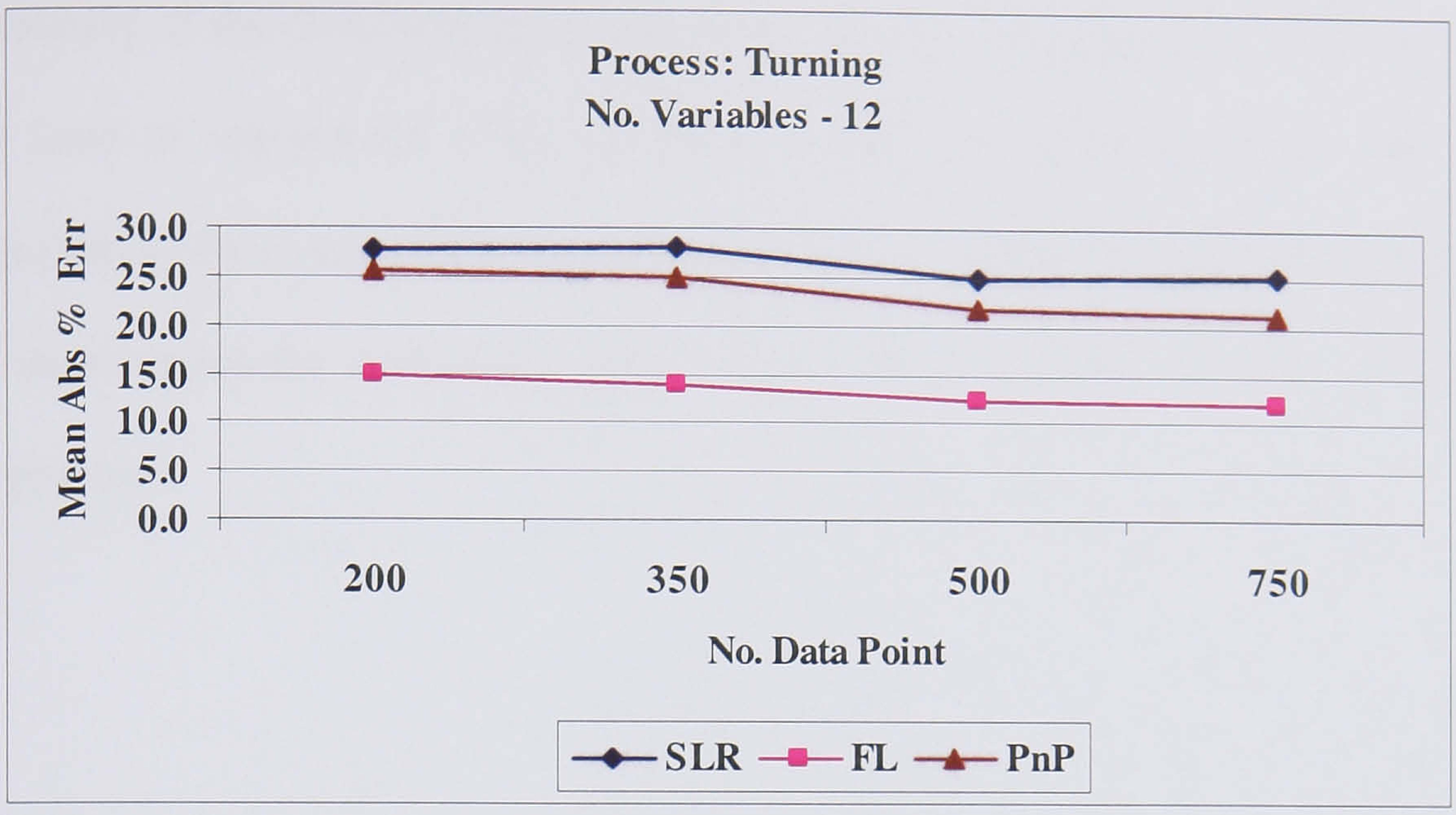
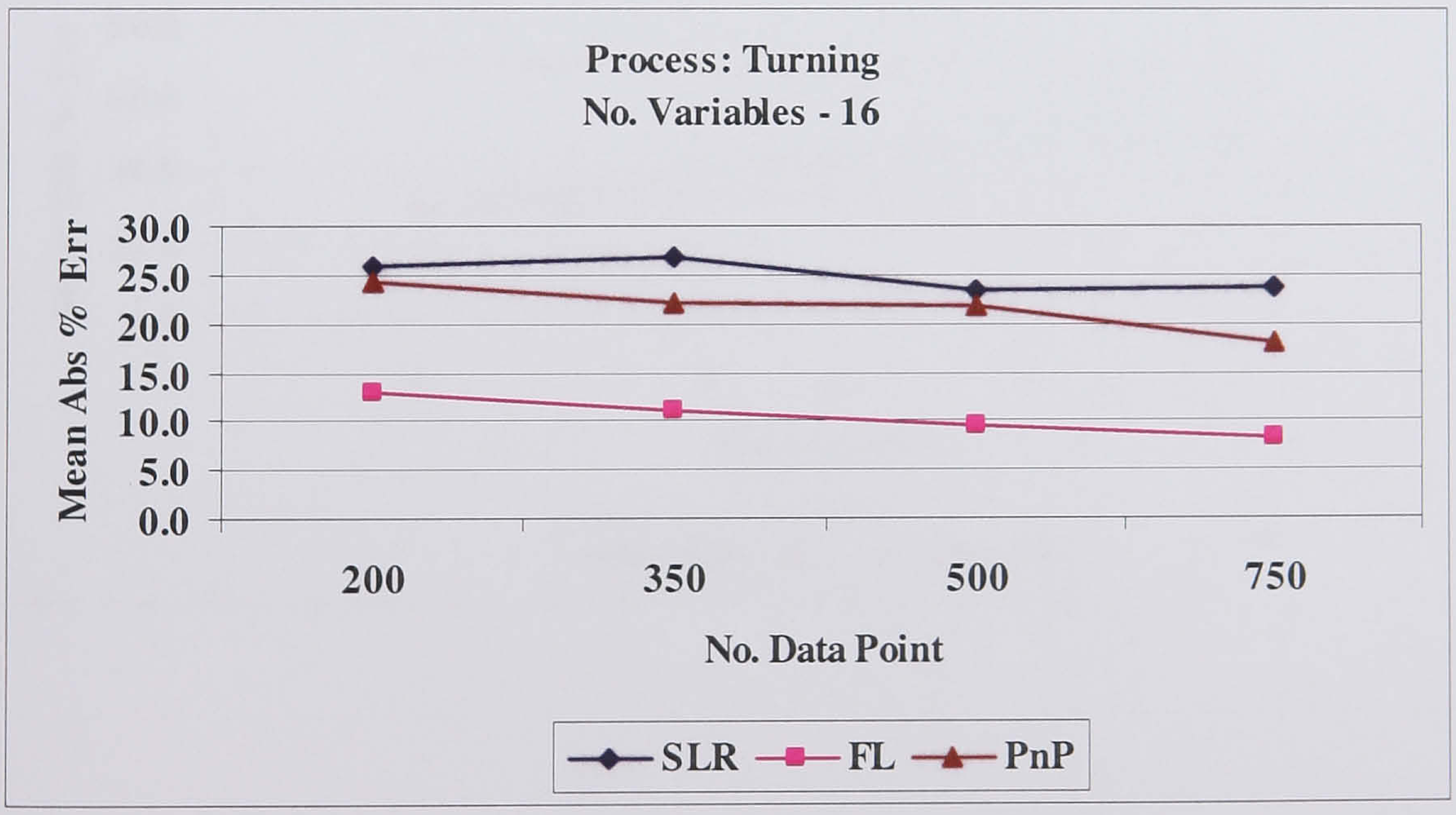


Figure 5. 18: Average Accuracy Vs Number of Data Points



5.3.2 Model Accuracy Vs Number of Variables

Experiments were carried out in order to identify the effect of number of process variables on the accuracy of the time estimating models developed. The maximum number of process variables used to explore the effect of these variables on model accuracy was 7 for the Automated Spray Paint and 16 for the Turning Process. Figures 5.19 to 5.22 provide results of these experiments for Automated Spray Paint process and Figures 5.23 to 5.26 for the Turning process.

Figure 5. 19: Average Accuracy Vs Number of Variables

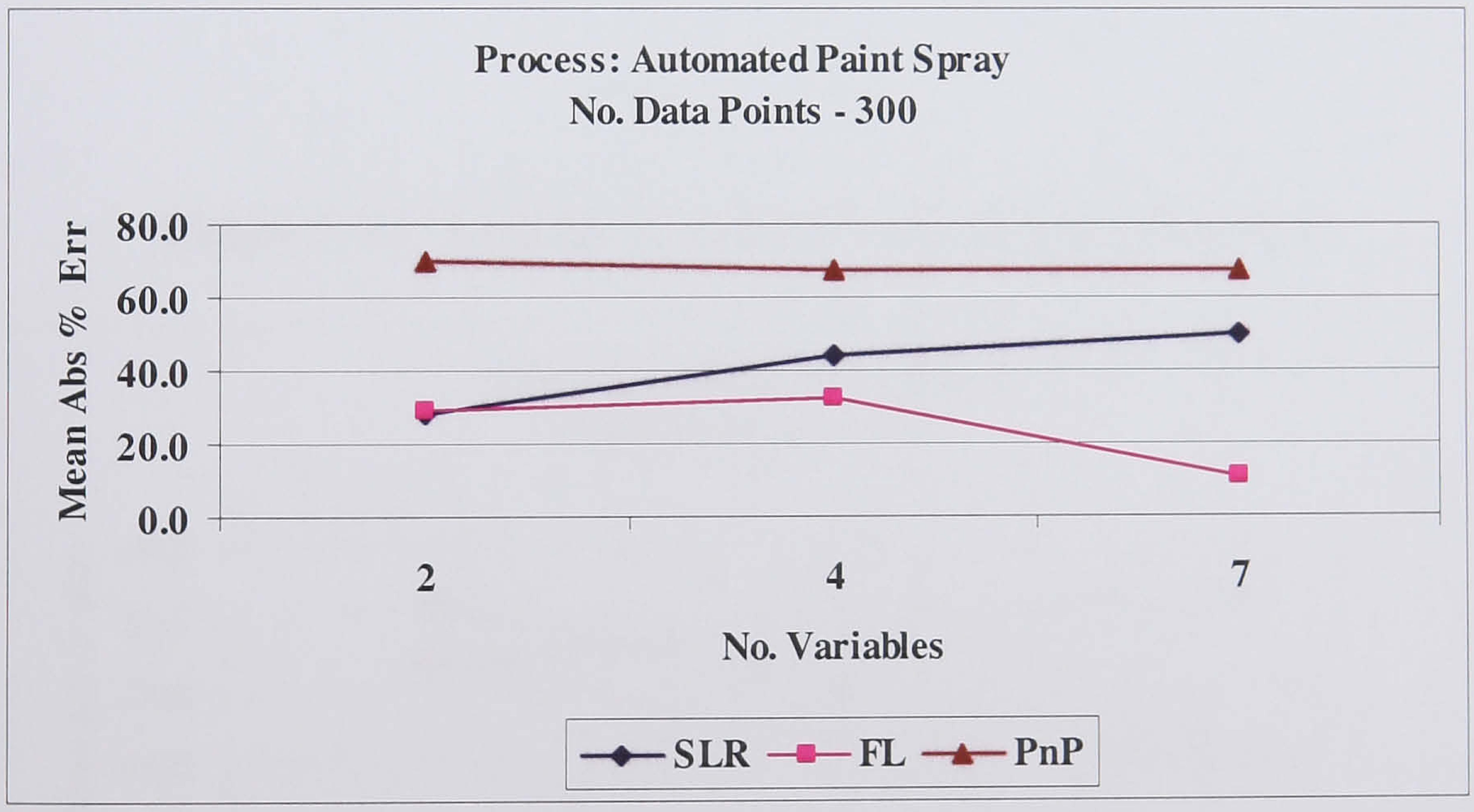


Figure 5. 20: Average Accuracy Vs Number of Variables

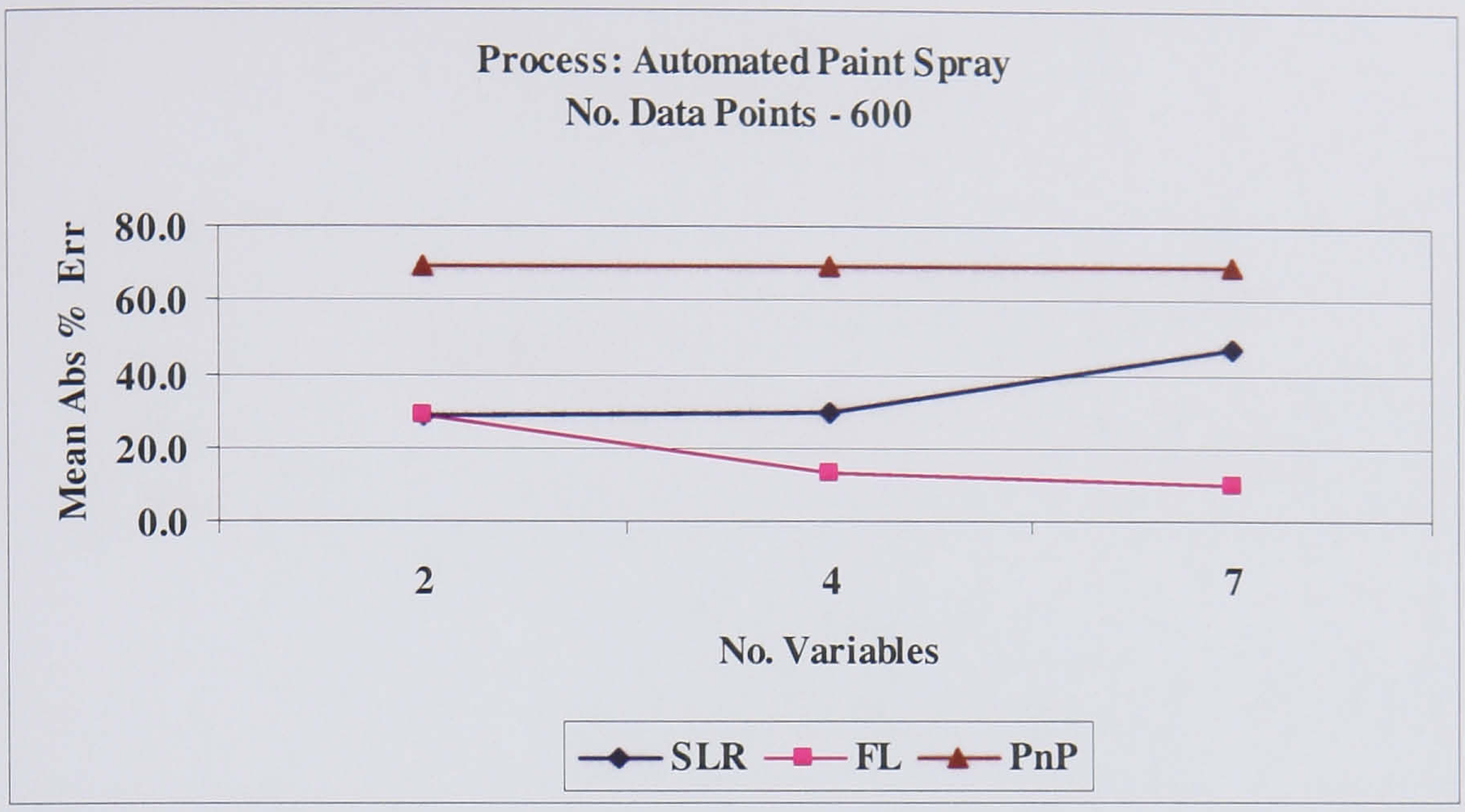


Figure 5. 21: Average Accuracy Vs Number of Variables

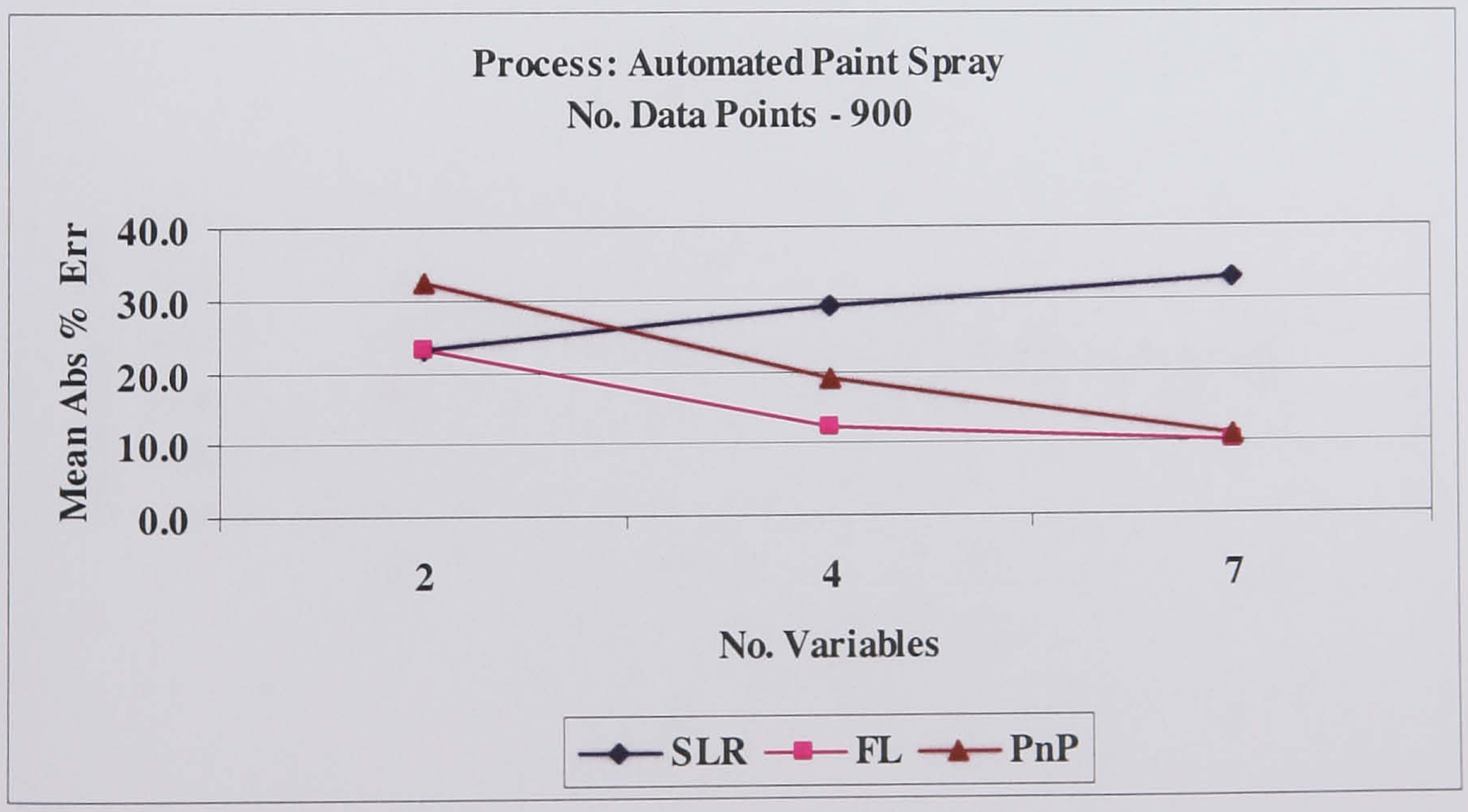


Figure 5. 22: Average Accuracy Vs Number of Variables

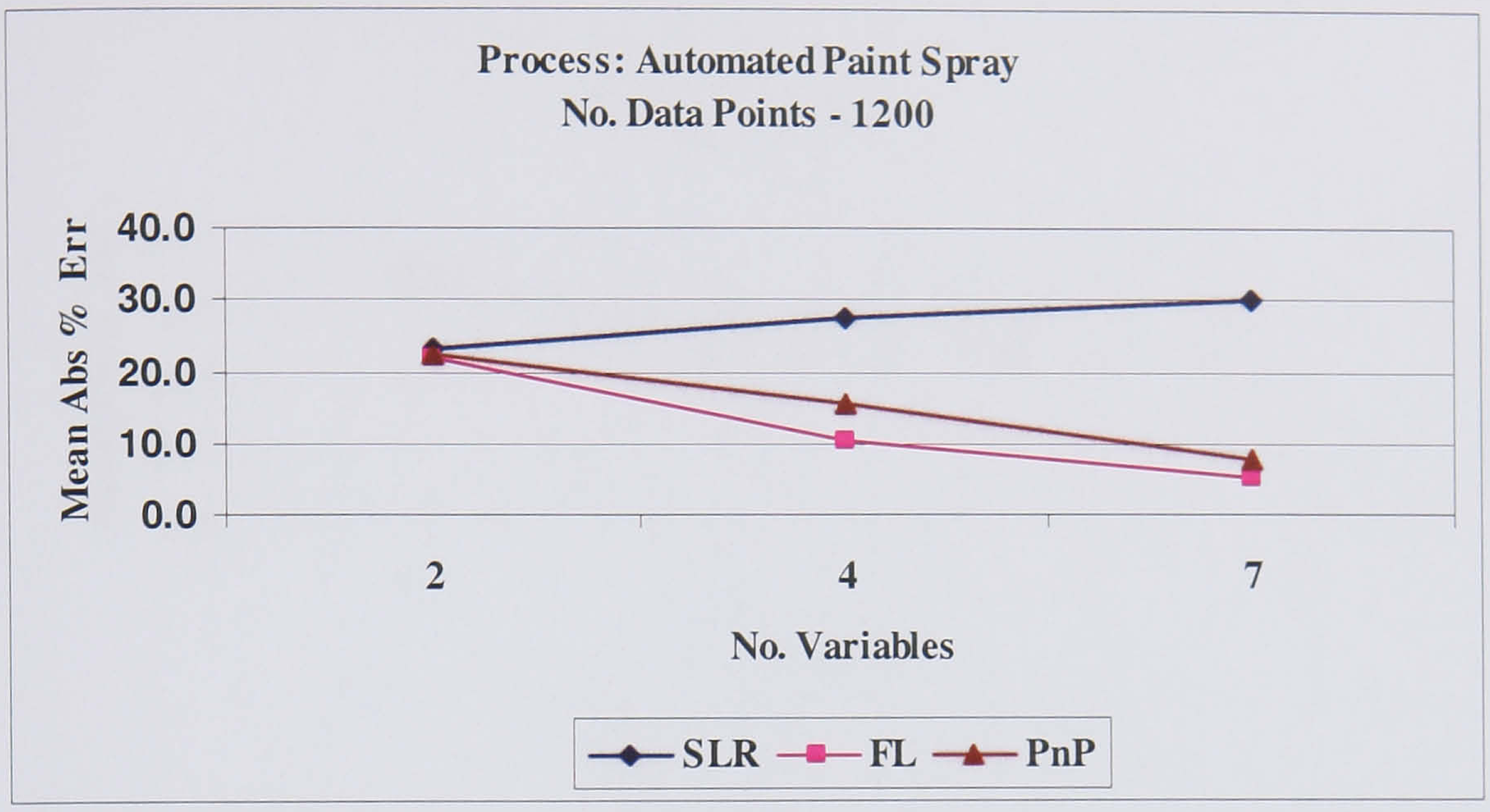


Figure 5. 23: Average Accuracy Vs Number of Variables

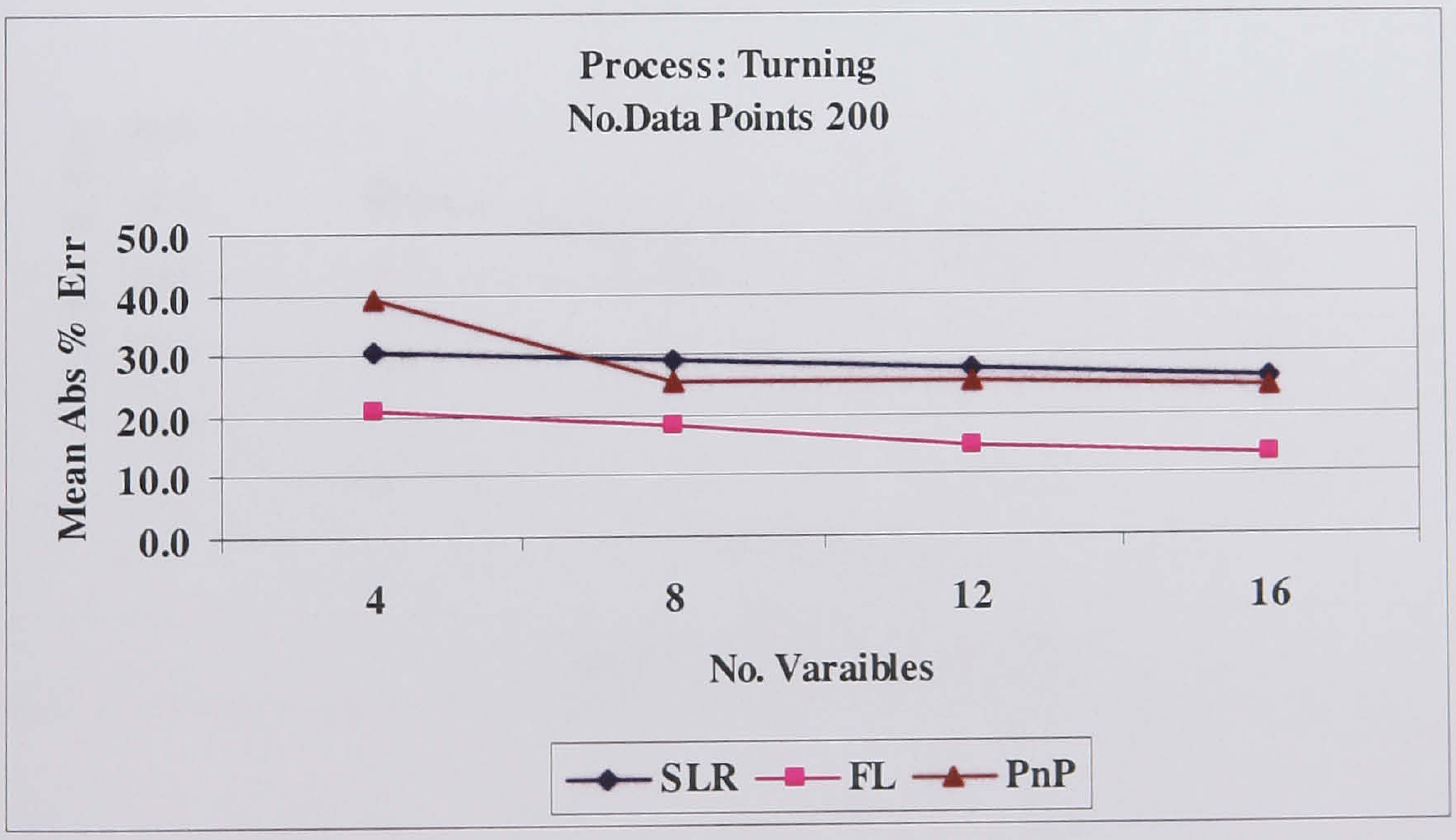


Figure 5. 24: Average Accuracy Vs Number of Variables

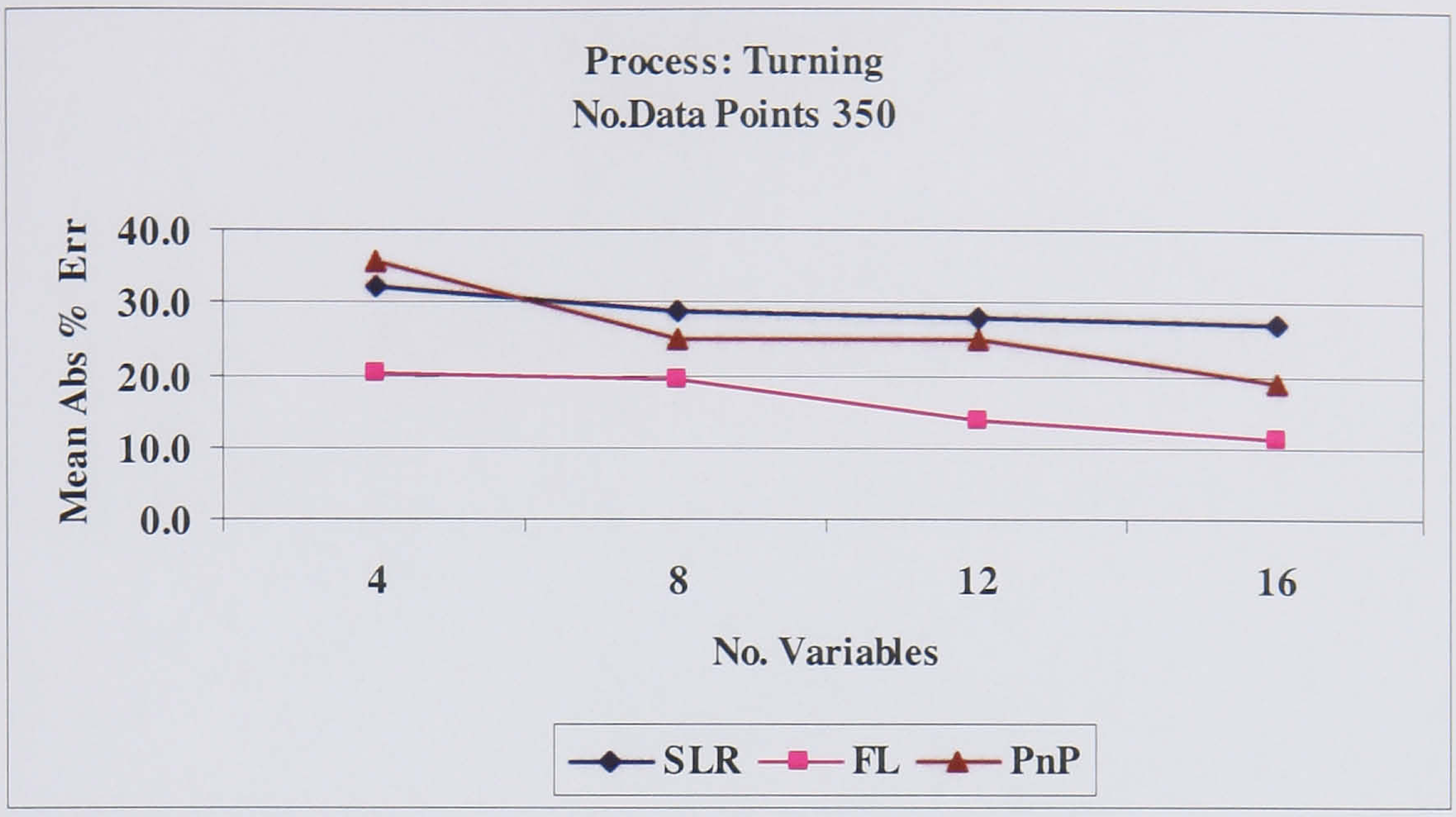


Figure 5. 25: Average Accuracy Vs Number of Variables

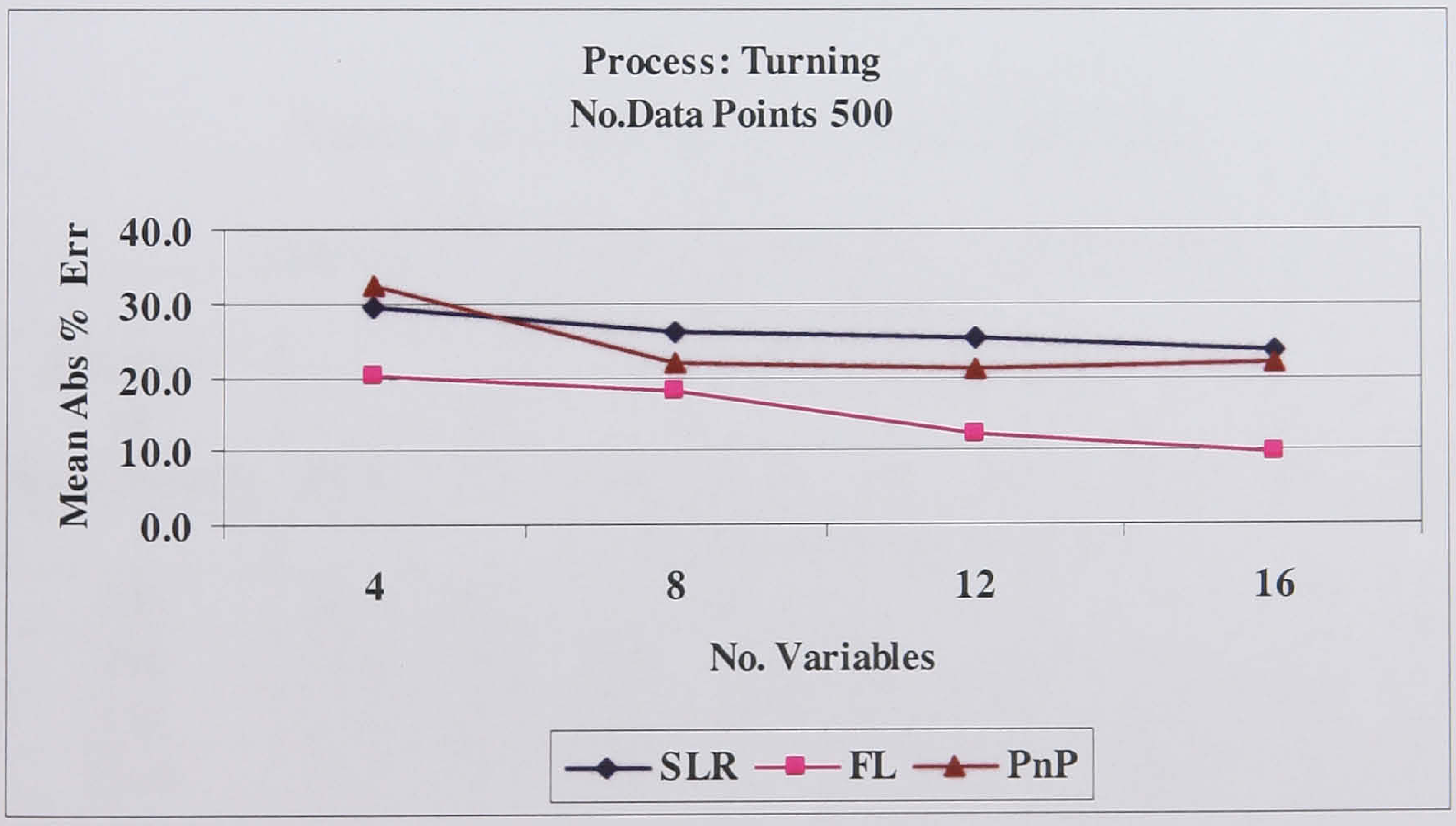
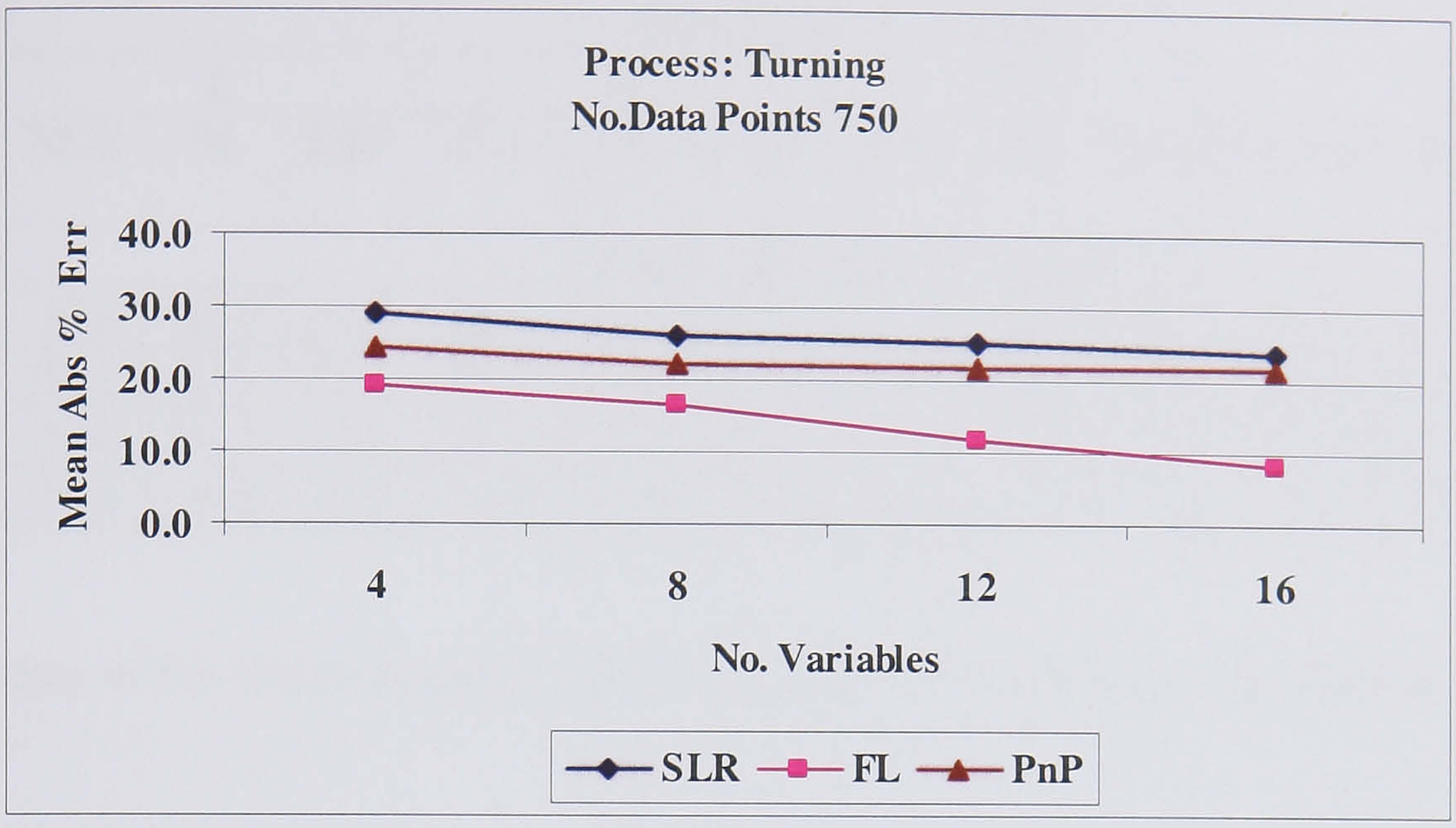


Figure 5. 26: Average Accuracy Vs Number of Variables



Tables 5.19 and Table 5.20 provide a summary of the combined effects of both "Number of Data Points" and "Number of Variables" on the estimating accuracy of the cost models developed.

Table 5. 21: Estimating Accuracy (MAPE)
vs
Number of Variables & Number of Data Points

Automated Spray Paint									
Number of Data Points	**Number of Variables**								
	2			**4**			**7**		
	SLR	**FL**	**PnP**	**SLR**	**FL**	**PnP**	**SLR**	**FL**	**PnP**
	Mean Absolute % Error								
300	28.3	29.5	70.2	44.4	32.1	67.6	50.1	10.9	67.6
600	28.6	28.6	69.4	29.9	13.3	69.4	46.7	10.3	69.4
900	23.3	23.3	32.6	28.9	12.2	19.1	32.8	9.9	10.9
1200	23.3	22.1	22.5	27.7	10.5	15.4	30.0	6.5	8.7

Table 5. 22: Estimating Accuracy (MAPE)
vs
Number of Variables & Number of Data Points

Turning												
Number of Data Points	**Number of Variables**											
	4			**8**			**12**			**16**		
	SLR	**FL**	**PnP**	**SLR**	**FL**	**PnP**	**SLR**	**FL**	**PnP**	**SLR**	**FL**	**PnP**
	Mean Absolute % Error											
200	30.8	20.7	39.1	29.0	18.4	25.5	27.7	14.8	25.5	26.0	13.1	24.3
350	32.2	20.1	35.5	29.0	19.4	24.9	28.0	13.8	24.9	27.0	11.2	22.1
500	29.6	20.1	32.6	26.3	18.2	22.1	25.1	12.4	21.8	23.5	9.6	21.8
750	29.2	19.0	24.3	26.3	16.6	22.2	25.3	11.9	21.1	23.7	8.1	18.1

The summary of the results listed in Table 5.21 and 5.22 are discussed in Section 6.3.1.

Chapter 6 Discussion

6.1 Introduction

Section 2.2.1 identifies the limitations of the existing CMD process in terms of its key process tasks and the characteristics of the resulting cost models. These process steps and key characteristics are now revisited to discuss how the use of VM and data mining can assist in overcoming the limitations of existing CMD methods.

6.2 Data Identification and Collection

6.2.1 Use of Subjective Expertise

A large proportion of total product and process life cycle costs is committed during the concept design stage (Roy 2001), i.e. they will be incurred through use of this concept. However, it is at this stage that current cost modelling methodologies rely heavily on the use of expert opinion because of lack of historical data. Here CER's are normally at their least levels of estimating accuracy and highest levels of estimating bias. The need for such high levels of expert opinion results in a CMD process that is both time and resource intensive and difficult to automate. Such expertise is also only available if historical data is available which presents a problem using traditional CMD processes when developing cost models for innovative / new products and processes.

Because of the high level of manual intervention traditional CMD processes tend to be highly iterative in nature and may often require many repetitions of the data identification, collection and analysis stages in order to develop a 'fit for purpose' cost model.

Traditional cost model development processes are, therefore, heavily reliant on the use of product and process experts, i.e:

a. At the data identification stage process expertise is required to identify the variables that effect costs and to identify valid data sources from where this data can be collected,
b. During the data collection stage manual input is required to select the most appropriate data collection method, to actually collect this data and to validate and verify the collected data,
c. During the data filtering stage data analysis expertise is needed to remove 'wrong' data items, to select an appropriate data analysis technique and to prepare the data for input in to the data analysis process,
d. At the data analysis stage expert opinion is needed to design and undertake the data analysis trials required to compare the effects, on estimating accuracy, of alternative predictor variables, to ensure the variables selected are the primary 'cost drivers' and to ensure the resulting model is of the highest accuracy levels,
e. identify the basic relationships, (linear or non-linear), between predictor variables and the dependent variable,
f. validate the model in terms of assessing if variable coefficients are feasible,
g. identify whether inter relationships exist between predictor variables, and
h. validate the causal effect of the predictor variables within the model is possible.

Sections 4.2 and 4.3, describe the requirements for the basic manufacturing process information that is required to build a VM process model. This indicates that the personnel involved in carrying out this task should have knowledge of the manufacturing process and the use of VM software in order to build an effective virtual model. Expertise required in areas such as the selection of cost drivers, the prioritisation of cost drivers and the selection of data sources can be greatly reduced over the levels required in existing cost modelling processes. Although the levels of expertise required for development of cost models can be greatly reduced, the use of VM methods does not allow it to be completely eliminated.

6.2.2 Levels of Data Detail and Quality

Through the literature research and the development of virtual process models in Section 4.2.2 and 4.2.3 it is has been identified that using VM models data at various levels of detail can be generated including at the process level, the assembly level and the operational activity level. In addition, VM models are capable of generating data that includes the main data items that make up cost models, i.e. product features, process features and process activities. The amount of data generated by the virtual process models, Table 6.2, indicates that there is no substantial limit to the maximum amount of data that can be generated. Product features, process features and process activity variables are represented in the model as is the effects of these variables on each other.

The accuracy of data generated from a VM model has been validated through Boothroyd's (2002) time estimation equation for vertical end milling as discussed in Section 6.3.3.1. Here also, the amount of data that can be generated from a single VM simulation run depends on the step size of the simulation. This ability to vary the amount of data to be generated is beneficial in ensuring flexibility in terms of the data analysis techniques used since the amount of data required to develop valid CERs varies for different data mining techniques. For example, from the observation of results in Table 5.19 and 5.20, in order to develop an accurate model i.e. with percentage error up to 20% the PnP method requires a minimum of 500 data points where as the FL algorithm requires only a minimum of 300 data points.

6.2.3 Cost Model Build Times

In terms of the time required to build cost models data collection is often the most time consuming task within the traditional CMD process. When collecting data from VM models three basic tasks are required, i.e. develop the VM model, carryout data generation experiments and collect the data generated by these experiments. An analysis of the alternative data generation methods, that are CS, DES, VM, MPM and VDG, was undertaken, (Table 3.4, 3.5 and 3.6) to determine the time required to undertake these basic tasks. It was identified from the results of these experiments that the time taken using VM is not considerably less than other data generation methods e.g. PMTS. However, from the experiments carried out it was found that within the CMD process the VM model building stage is the only significant time consuming task of these three basic steps. VM models also have the benefit that once built they can be reused, with the required modifications, at later stages of process development, hence, saving cost modelling time and resources.

The benefits described above need to be considered when deciding to adopt a VM data generation approach since the overall times required to produce a cost model may not be reduced. Here Table 6.1 shows the times required to develop virtual process models that are suitable for generating data from which cost models may be developed. These model building times include two to four weeks for training in use of the VM software, (i.e. the Machining and Automated Paint Spray modules) and approximately one week per person to develop the models. These time values represent the actual times taken to build the models and could equal the data collection times that arise during the traditional CMD process. In addition, Sections 4.2.1 to 4.2.4 provide details of the data sources and resources required to build the models, which indicate the areas that process and/or product expertise is still required within the CMD process.

Table 6. 1: Time to build virtual process models

Manufacturing Process	No. Models	Time to Build VM Models
Vertical End Milling (Figures 5.1 to 5.3)	3	240 (hours)
Automated Paint Spraying (Figures 5.4 to 5.7)	4	160 (hours)

Traditional CMD processes, Section 2.2.1, consist of numerous iterations of the basic tasks of data identification, data collection and data analysis. The essential aim of these iterations is to identify and collect sufficient historical data and to analyse this data such that valid and accurate cost estimating relationships can be identified. These iterations involve manual decision making to determine the direction of future iterations, e.g. what additional data sources may be used.

The methods developed by Delgado, McNeill, Stockton (2002) help to improve the structure of the CMD process but still rely heavily on the use of process, product and cost modeling expert opinion. Their methodologies merely enable the CMD process to be undertaken by multi-skilled teams to reduce the number of complete iterations, i.e. data identification, data collection and data analysis, involved in the CMD process and assists in improved identifications of the 'cost driver' variables.

Using VM and data mining techniques the number of iterations required to produce valid and accurate CER, may not be greatly reduced when compared with traditional CMD processes, but the time taken to undertake each iteration would be greatly reduced. Hence, leading to much reduced CMD times.

6.2.4 Data Generation

The following process was developed during the research for generating data using VM models, i.e:

i. Collect process description, i.e. process features, product features and process activities.
ii. Identify the process constraints, e.g. feasible process variables such as surface speed, ratio of tool diameter and depth of cut for milling.
iii. Collect maximum and minimum values of product design features.

iv. Develop virtual process models using the process description and product information. Ensure that all relevant product features, process features and process activity variables are represented in the model as is the effect of these variables on each other.

v. Test the validity of the virtual process model through observation. Here VM model's dynamic operational behavior is displayed graphically, i.e. as discrete activities occur during the model run time. For example, the individual cutting actions of a tool on a target work piece are shown graphically as they occur over period of time.

vi. Design experimental trials using Taguchi Orthogonal Arrays (OAs).

vii. Run OA trials to generate data.

vii. Collect the data in a format, which is compatible with the data analysis process.

6.2.5 Benefits of Data Generation

When collecting data it is not possible to 'control' the values of the predictor variables that are collected. Hence, these values may:

a. be biased towards low, medium or high values,
b. not contain minimum and / or maximum values of the predictor found in practice, and
c. contain insufficient number of data points when compared with number of predictor variables.

In addition, when verifying and validating data from existing data sources often the assumptions and/or constraints under which the data was collected may be unknown. Hence making it difficult to under take such validation and verification.

Developing a data generation method overcomes problems with the lack of suitable historical data sources from which to collect data. It also overcomes the need to initially identify the data to be collected, the identification of relevant data sources, the selection of a suitable data collection method, the collection of the data from its sources and the validation and verification of the collected data. Since each of these tasks requires significant manual input use of a data generation tool would automatically remove this from the CMD process. It does, however, require the development of suitable virtual models which include in them the basic cost data types such as product features, process features and process activities and this suitability of the models is judged by the estimator using the model and undertaking virtual trials with these models in order to generate data. Development of models necessitates undertaking validation trials and verification trials of the resulting models. Part of these would involve ensuring that VM models were representative of the product or process to be developed and contained all process features and activities involved. In terms of product features the trials undertaken to generate data ensured that the range of virtual products used represented the full range of product features and their parameter ranges that would be found in practice, i.e. as shown in Section 4.2.2 and 4.2.4 for the Milling and Automated Paint Spray VM models.

In addition, use of VM models to generate data, enables use of the Taguchi methodology to design series of experiments which maximise the effectiveness of the data generation process. Employing the Taguchi methodology achieves this aim and in addition reduces the experimental design process (Rowlands 2000) by minimising the number of experiments necessary to generate data, i.e. as shown in Tables 4.8 and 4.10. When developing cost models the accuracy of data and amount of data required are key concerns. Minimising the amount of data required through use of Taguchi OA's allows greater emphasis to be placed on ensuring the accuracy of the data.

It was also found that care needed to be taken when defining process variable levels for OA's since the possibility could exist of designing an experiment that was not feasible in practice. For example, an OA experiment that required a depth of cut of 75 mm and tool diameter of 6.35mm is not a feasible combination since in practice the tool could be expected to fail during the cutting process. Therefore, users should provide values in such a way that OA experiments provide valid results. Manufacturing process expertise is, therefore, required in order to check if OA experiments are possible in real life situations.
It also provides the opportunity of using DOE methods, limited by feasible values, to design virtual trials that:

a. Minimises number of virtual trials required,
b. Provides structured method of ensuring min/max values of each predictor variable is represented in the trials, and
c. Attempts to remove the effect of other variables on individual variables effects.

6.2.6 Time to carry out Data Generation Experiments

Tables 3.4b and 3.4c, list the requirements of data types for various data generation methods. Tables 4.8 and 4.10 provide the list of experiments carried out for the purposes of data generation. It was found that the time required to carry out these experiments was approximately 160 man hours. The amount of data, i.e. number of data points, that has been generated for Vertical End Milling and Automated Paint Spraying processes are listed in Table 6.2.

Table 6. 2: Number of Data Points generated from Virtual Process Models

Process	Number of Data points
Vertical End Milling	1679
Automated Paint Spray	1188

From the experiments carried out it was found that in order to generate data the time required to carry out the experiments was dependent on such factors as:

a. complexity of the process being modelled,

b. the amount of data required from the experiments, and

c. the type of data required.

6.3 Data Analysis

The various data analysis methods used to develop cost models were identified and compared in terms of their ability to function under conditions where data was scarce, i.e. at the concept product design. A range of data analysis methods have been used to develop cost models at the conceptual stage of product design including, probabilistic estimation

(Sonmez 2005), parametric (Stewart 1987; Mileham 1993), expert opinions/subjective judgment (Roy 2002), probabilistic cost estimating (Touran 1993), Artificial Neural Networks (Wang 200; Bode 1997) and fuzzy logic (Jahan-Shahi 1999).

However, from the analysis in Section 3.6, it was evident that data mining seemed to be a more appropriate method when compared with other data analysis tools, i.e. data mining tools are able to:

a) identify and prioritize cost drivers prior to the data analysis stage,
b) deal with both linear and non linear relationships simultaneously,
c) provide a predictive model not a "black box",
d) deal with large numbers of data points and variables, and
e) provide multiple outputs.

6.3.1 Estimating Accuracy: Effect of Number of Data Points and Predictor Variables

The estimating accuracy of CERs developed using data mining algorithms has been found to be strongly dependent on the number of data points used to derive the models. This has been demonstrated in the experimental results provided in Figures 5.10 to 5.24. Here, Figures 5.10 to 5.12 show the effect of Number of Data Points and Figures 5.17 to 5.20 show the effects of Number of Variables on the estimating accuracy of the Automated Spray Paint process and indicates that:

a. As the number of data points increase the estimating accuracy of the models developed using each of the algorithms, i.e. SLR, FL and PnP, also increases,
b. The highest level of accuracy is the SLR derived model, which is 23.3%, using 1200 data points. These results indicate the presence of non-linear relationships between predictor and independent variables. The accuracy of the SLR model tends to decrease with increasing numbers of predictor variables. This indicates that more non-linear variables are being included in models and hence decreasing their "linear" accuracy,
c. The FL derived model provides a consistently high level of estimating accuracy across a wide range of process times with a maximum accuracy of 5.1%. FL also provides a consistent increase in estimating accuracy as the numbers of predictor variables increase, i.e. as the number of non-linear variables and data set sizes increase estimating accuracy using the FL derived models increases.
d. The PnP derived model requires a threshold level of data points, i.e. 600, before its estimating accuracy begins to improve. This improvement ends at approximately 900 data points. This indicates that the estimating accuracy of the PnP derived model was low for fewer data points, i.e. 600. Therefore use of the PnP method to generate cost models is best suitable when there are high numbers of data points and/or variables.

Figures 5.13 to 5.16 and Figures 5.21 to 5.24 show the effect of Number of Data Points and Number of Variables on the estimating accuracy of the Turning process and indicate that:

a. in all cases as the sample size of the data set and number of predictor variables increases the estimating of resulting CERs accuracy also tends to increase,

b. SLR derived models provide a consistent increase in the estimating accuracy with increasing numbers of data points and numbers of predictor variables with the maximum accuracy of SLR being 23.7%.

c. FL derived models provide the maximum levels of estimating accuracy, i.e. 8.1%, using 750 data points and 16 variables.

d. PnP derived models again provide a consistent increase in estimating accuracy, as found with the use of SLR and FL, and yields a maximum estimating accuracy of 18.1% using 750 data points and 16 variables.

Tables 5.23 and 5.24 provide a summary of the results obtained for Automated Paint Spraying and Turning. From these results, general rules for using each of the data mining algorithms, (i.e. SLR, FL and PnP), can be derived and are provided in Table 6.3.

Table 6. 3: Rules for Using Data Mining Algorithms

Data Mining Method	No. of Data Points		No. of Variables		Relationships	
	≤ 400	≥400	≤3	≥4	Linear	Non-Linear
FD		√		√	√	√
SLR	√		√	√	√	
FL		√	√	√		√
PnP		√		√	√	√
FD & SLR	√	√	√	√	√	
FD & FL	√	√	√	√		√
FD & PnP		√		√	√	√

6.3.2 Analysis of Results

Sections 5.1.1.1, 5.1.2.1 and 5.1.3.1 show the results obtained from the preliminary analysis of the Find Dependencies algorithm. The purpose of this analysis was to identify:

a. the potential dependencies that existed between predictor and dependent variables, i.e. the identification of process cost drivers, and

b. if "exceptional" values existed in the data which if excluded by the FD algorithm may be helpful in development of the CER, i.e., lead to higher levels of estimating accuracy.

6.3.2.1 Vertical End Milling

Table 5.1 provides the results obtained using the FD algorithm for the Vertical End Milling process, i.e.:

a. the most influential independent variables effecting process times are Ft, V_f and L_c ,
b. the "P" value of the dependencies is "0", hence the prediction of the FD algorithm is accurate in terms of its selection of influential independent variables,
c. of the total data set of 1679 data points, the FD algorithms has split the data set into two data sub-sets i.e. DS_1 and DS_2. DS_1 has 1433 data points and the FD table for this data set is provided in Appendix 6.1A. From the FD Table 6.1A it can be observed that DS_1 has the following characteristics:

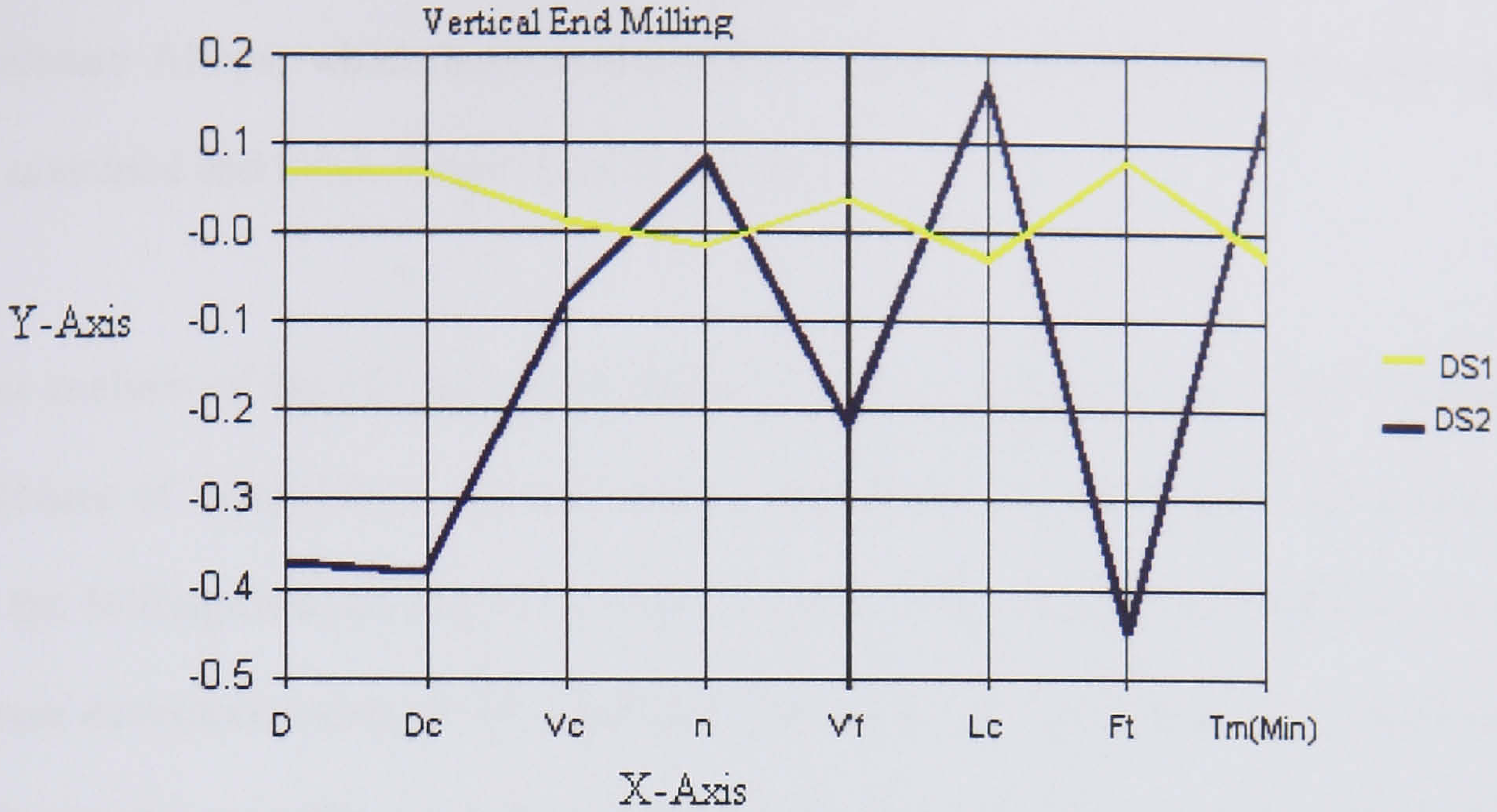
    a. for constant machining length, as feed rate increases the machining time decreases,

b. for constant feed rate, as machining length increases the machining time increases, and

c. values with lower feed rate and small tool diameters were not found in this distribution pattern.

In order to identify why a proportion of DS_2 values were classed as 'exceptions', i.e. not included in DS_1, both data sub-sets were plotted on an HiLo chart as shown in Figure 6.1. Here, the X-axis represents the process variables and the Y-axis represents the values of these variables in their "normalized by dispersion form", i.e. in the range -1 to 1. It is evident from Figure 6.1 that DS_2 has values that are distributed over a wider range than DS_1 values. This chart also reveals that DS_2 machining times are more influenced by lower values of the depth of cut, feed rate and tool diameter predictor variables. It is this combination that makes them exceptions when compared to the complementary data set DS_1.

Figure 6. 1: Milling HILO Chart



X-axis: Milling variables, Y-axis: values of variables "normalised by dispersion"

In order to illustrate the effect on machining times of these independent process variables in DS_2 Figures 6.1A and 6.1B have been provided in Appendix 6.1A. These figures reveal that the values contained in DS_2 possess:

a. lower feed rates,
b. lower depths of cut,
c. low feeds per tooth,
d. high machining lengths, and
e. higher machining times.

These figures also reveal that the majority of the data points in this data set belong to High Temperature Alloys, which have comparatively lower feed rates then the other material types examined and hence longer machining times.

Further analysis of the FD results for Milling has been carried out in order to identify the significance of using FD as a preprocessing algorithm, i.e. experiments were carried out using the Milling data set and its two data sub-sets. The models resulting from these data sets were developed using the FL algorithm. The estimates generated using these resulting models are shown in Figures 6.2 and 6.3, which show that DS_2 and DS_1 are similar when the tool diameter and depth of cut is less. However, it can be seen in Figure 6.3 that DS_2 curve tends to move away from the data points with increasing depth of cut and tool diameter.

Figure 6. 2: Comparing CERs for Milling DS_1 and DS_2 (i)

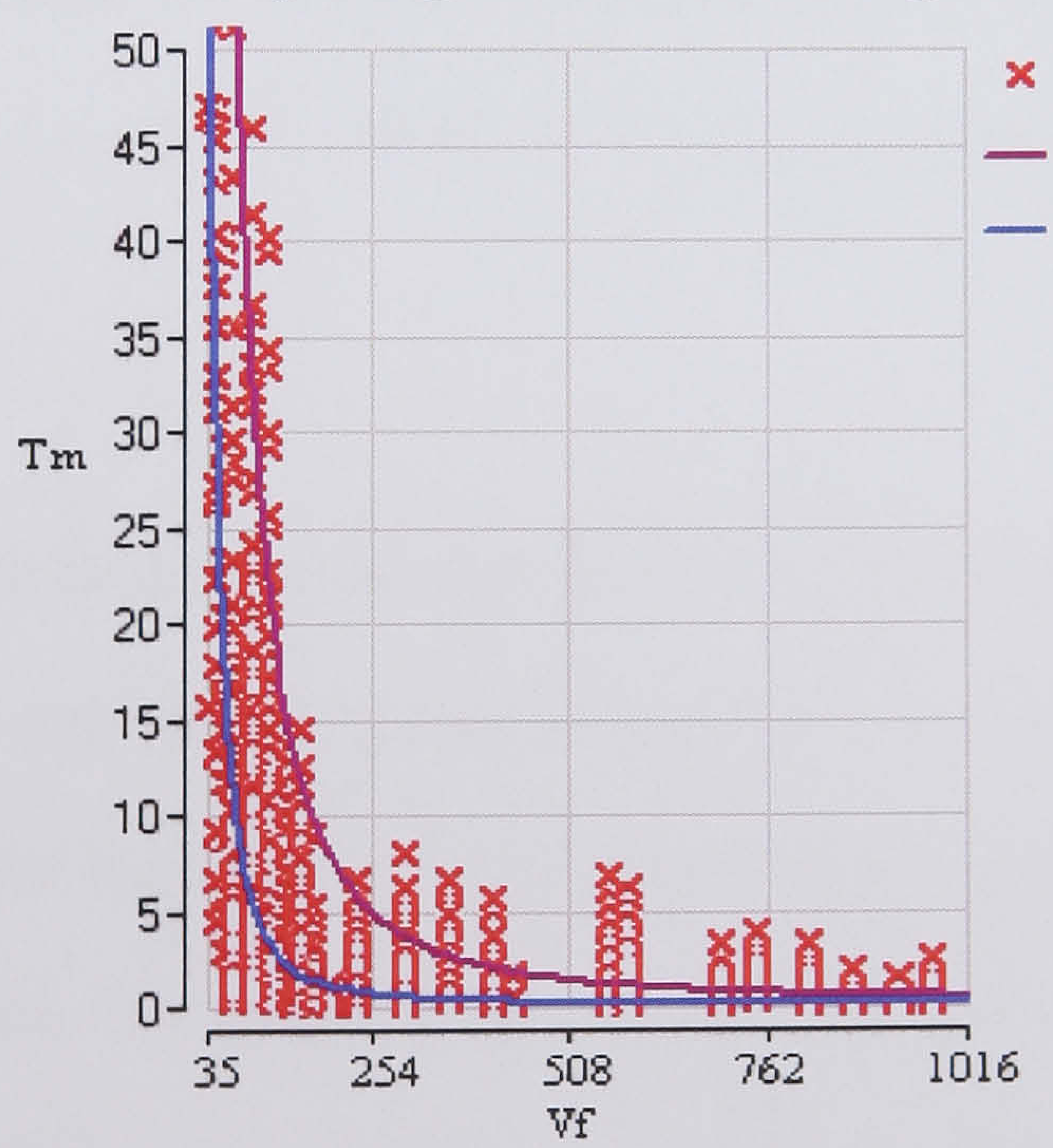
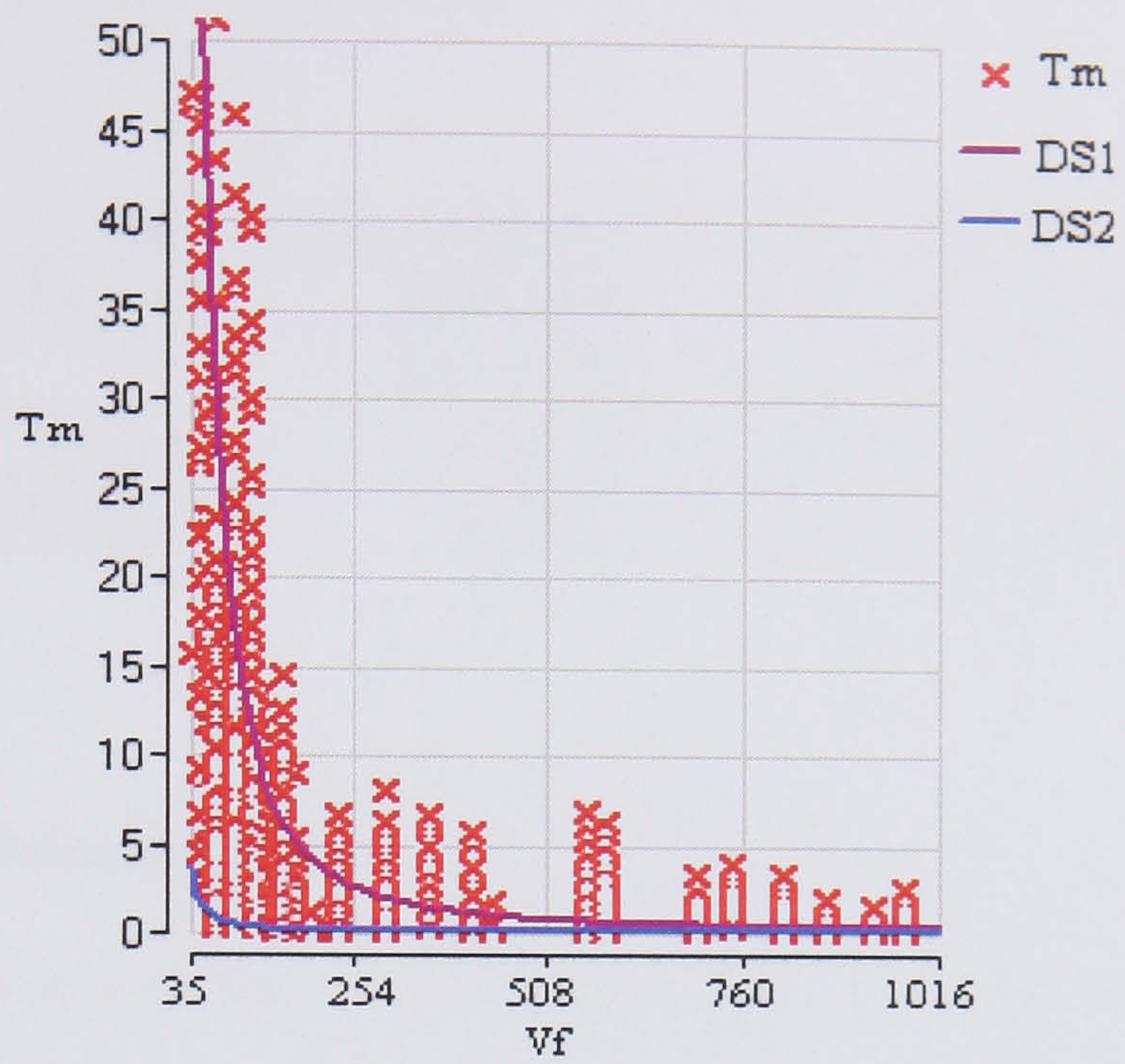


Figure 6. 3: Comparing CERs for Milling DS_1 and DS_2 (ii)



The results shown in Figure 6.2 and 6.3 illustrate that the estimating models derived from DS_2 are influenced by the tool diameter and depth of cut, only after a certain value of tool diameter has been exceded, i.e. 14.77mm. Where as for estimating models derived from the DS_1 data set, the Ft, Lc and V_f values are found to possess the highest influence on machining times.

The estimates generated from the main dataset and its two data sub-sets (DS_1 and DS_2) were compared in terms of estimating accuracy levels (i.e. Table 6.5). These results indicated that there is a significant improvement in the estimating accuracy through use of the Find Dependencies algorithm to identify potential data sub-sets. However, further analysis indicates that these higher accuracy values are mainly due to the existence of low process time values. When the 'percentage errors' of these values is measured the resulting error is magnified hence resulting in higher average error values.

Table 6. 4:Accuracies of Milling Models

Process	**Main DS**	**DS1**	**DS2**
	Before FD	After FD	
Milling	1679	1433	246
Accuracy (%)	93.60	31.88	69.57

6.3.2.2 Automated Paint Spraying

Table 5.10 provides a summary of the results obtained using the Find Dependencies algorithm to analyse the Automated Paint Spray data set, i.e.:

a. the most influential independent variables effecting process times are G_s and P_{tl},

b. the "P" value of the dependencies is "0", hence the prediction of the FD algorithm is accurate in terms of its selection of influential variables,

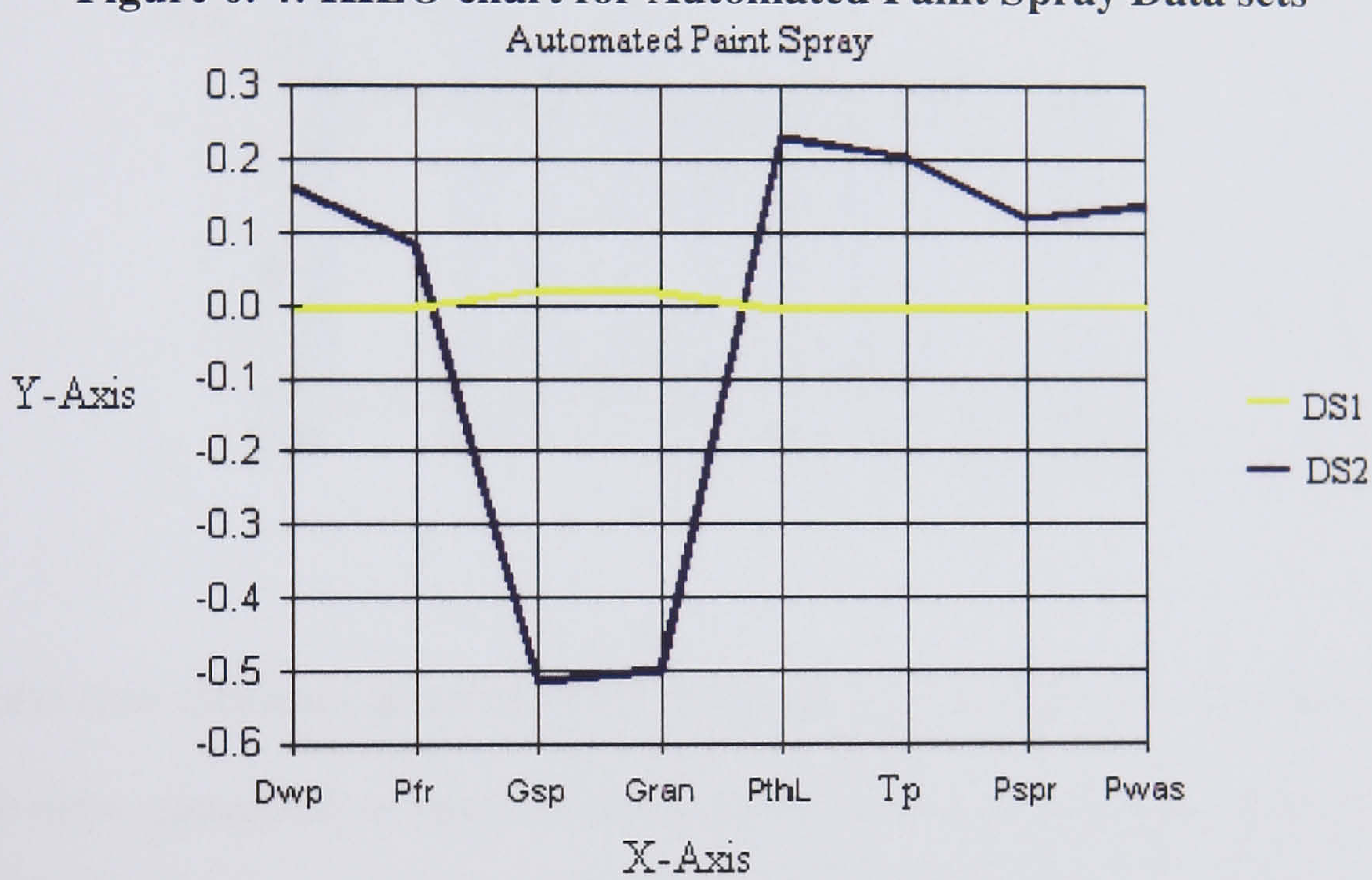
c. from the main data set of 1189 data points, FD has identified two data sub-sets, i.e. DS_1 and DS_2. DS_1 has 1148 data points and the FD Table 6.2A for this data set is provided in Appendix 6.2A. It is evident from Table 6.2A, DS_1 has the following characteristics:

a. for constant paint gun speed, by increasing length of the painting area the time for painting is increased, and

b. for constant length of paint area, by decreasing paint gun speed the process time also increases.

In order to identify why a proportion of DS_2 values were classed as exceptions, i.e. not included in DS_1, both data sub-sets were plotted using an HiLo chart as shown in Figure 6.4. "Thermal Charts", and also provided in Figures 6.2A and 6.2B, in Appendix 6.2A. These figures indicate that a small number of data points in DS_2 behave in an opposite manner to that of the data points in DS_1. It was this difference in behavior that caused them to be excluded from the dependencies identified by the FD Algorithm.

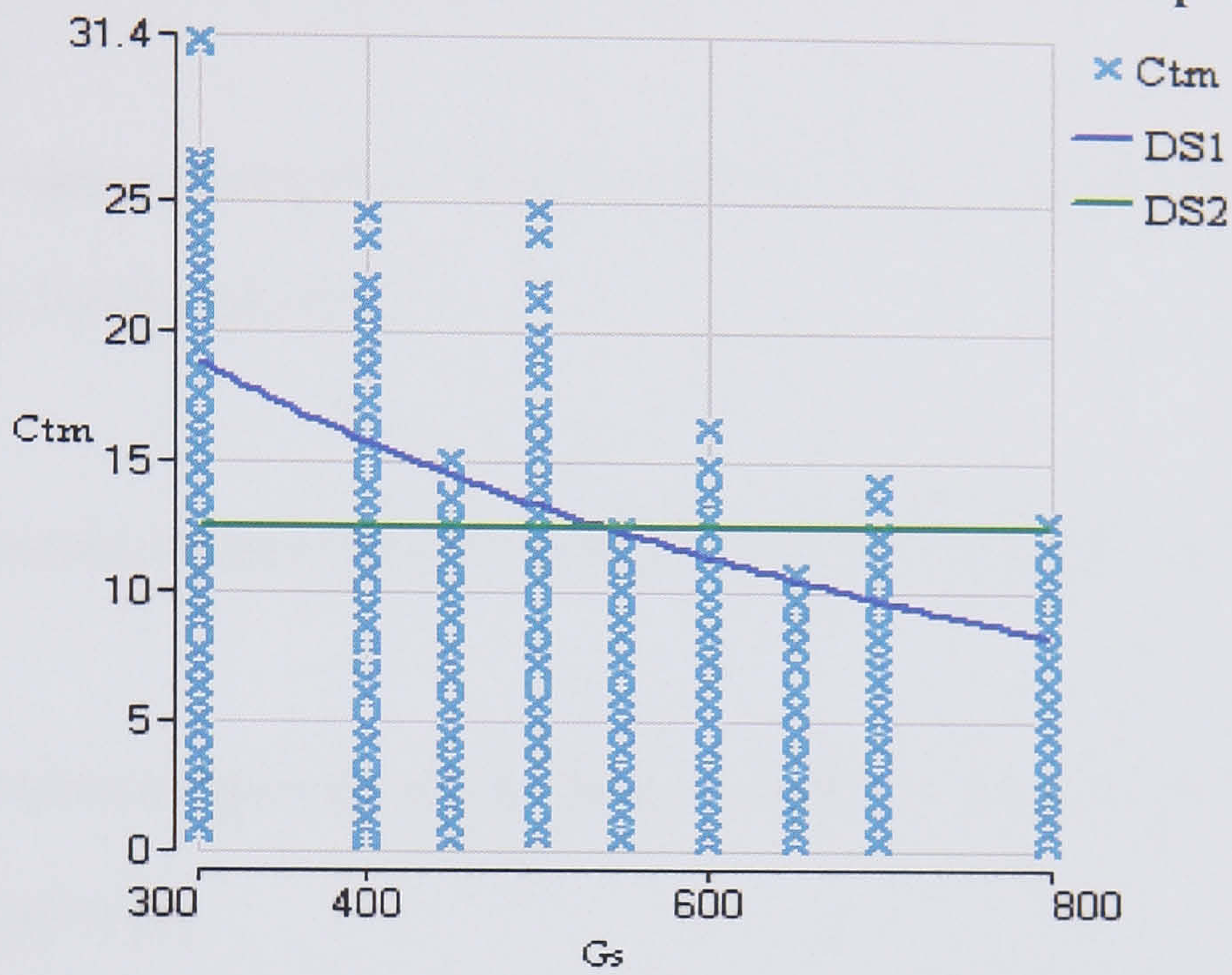
Figure 6. 4: HILO chart for Automated Paint Spray Data sets



X-axis: represents Automated Paint Spray process variables
Y-axis: represents the values of these variables in "normalised by dispersion" form

Further analysis of the FD results for Automated Paint Spraying was undertaken in order to identify the significance of using FD as a preprocessing algorithm, i.e. experiments were carried out using the automated paint spray data set and its two data sub-sets. The resulting process time estimating models from these data sets were developed using the FL algorithm and are illustrated in Figure 6.5. This figure indicates that the variables in DS_2 tend to behave linearly with time when compared to those in DS_1. This behavior of data points in DS_2 makes them exceptional, i.e. they are therefore not included in DS_1.

Figure 6. 5: Comparing CERs for Automated Paint Spray



The process time estimates generated from the main dataset and its two data sub-sets (DS_1 and DS_2) were compared in terms of estimating accuracy levels (i.e. Table 6.6). These results indicate that there is no marked increase in the estimating accuracies of the models developed with or without using FD. However, the model developed from DS_2 was found to have lower levels of estimating accuracy than that developed from DS_1 and process times are more influenced by the variable P_{fr}, i.e. "paint flow rate". The process time estimating model developed from DS_2 is not applicable when paint flow rate values are in the range of 436.5 to 508 cc/min.

Table 6. 5: Accuracies of Automated Paint Spray FL Models

	Main DS	**DS_1**	**DS_2**
	Before FD	After FD	
No. Data Points	1189	1148	41
Accuracy (%)	6.51	5.58	22.50

6.3.2.3 Turning

Table 5.10 provides a summary of the results obtained using the Find Dependencies algorithm for the Turning process, i.e.:

a. the most influential independent variables effecting turning process times are n_t , B_s , t_{ln} and t_{pt},

b. the "P" value of their dependencies is less then 0.00001 indicating that the dependencies found are highly relevant,

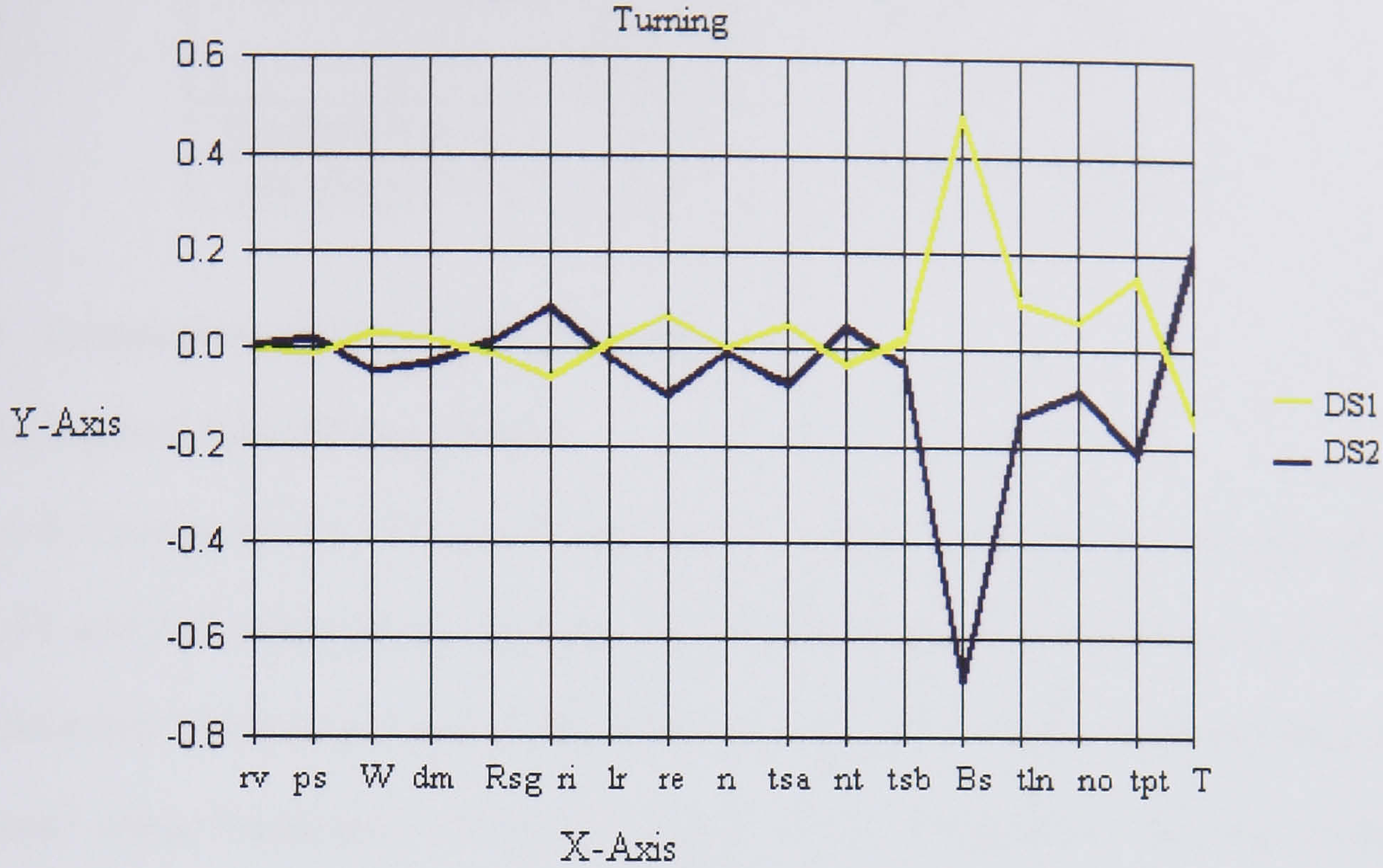
c. from the data set of 749 data points, the FD algorithm identified two data sub-sets, i.e. DS_1 and DS_2. DS_1 has 440 data points and the FD table for this data set is provided in Appendix 6.3. It is evident from these results that DS_1 has the following characteristics:

a. batch size was found to be the highest influencing factor on turning process times, and

b. DS_1 data points have batch sizes greater than 54 for all the values of n_t, t_{ln} and t_{pt}.

In order to identify why DS_2 values were classed as 'exceptions', i.e. not included in DS_1, both data sub-sets were analysed using HiLo charts (i.e. Figure 6.6) and "Thermals" charts, i.e. Figures 6.3A and 6.3B, in Appendix 6.3A. These figures reveal that both of these data sub-sets are mirror images of each other. Hence, if the values of one variable are lower in DS_1 then the values of same variable are higher in DS_2. DS_1 is characterised by data points with higher batch sizes where as DS_2 is characterized by data points with lower batch sizes.

Figure 6. 6: HILO chart for Turning Data Sub-Sets



X-axis: represents Turning process variables
Y-axis: represents the values of these variables in "normalised by dispersion" form

Further analysis of the FD results for Turning has been carried out in order to identify the significance of using FD as a preprocessing algorithm, i.e. experiments were carried out using the Turning data set and its two data sub-sets and the process time estimating models resulting from analysis of these data sets, i.e. using the FL algorithm were analysed. The estimates generated from the main dataset and its two data sub-sets (DS_1 and DS_2) were compared in terms of estimating accuracy levels (Table 6.7). These results indicate that there is no significant improvement in the estimating accuracy through application of the Find Dependencies algorithm.

Table 6. 6: Accuracies of Turning FL models

	Main DS	DS_1	DS_2
	Before FD	After FD	
No. Data Points	749	440	309
Accuracy (%)	8.13	7.94	13.25

6.3.3 Comparisons of Predictive Models-I

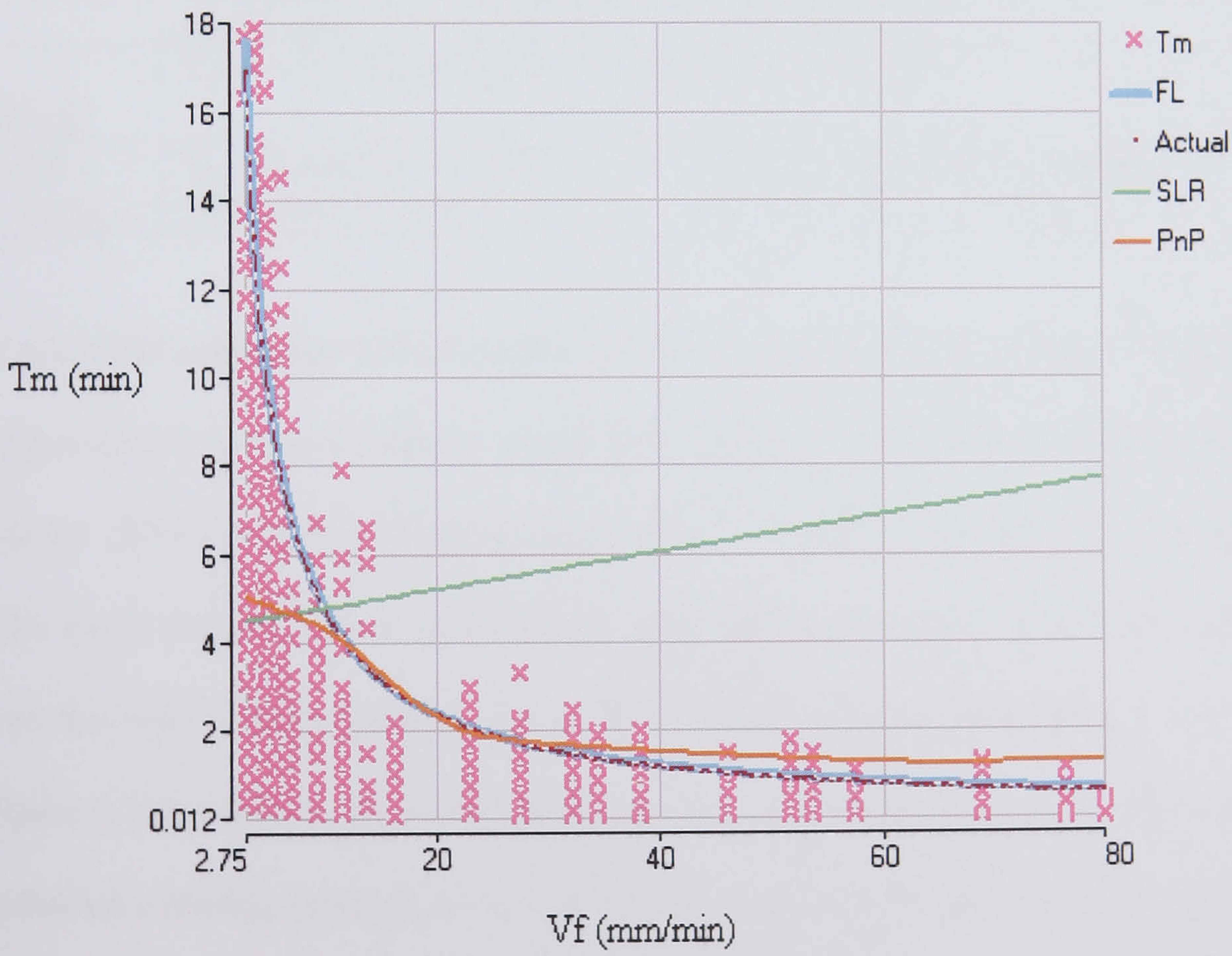
6.3.3.1 Vertical End Milling Models

Figure 6.7 compares the CERs developed for the Vertical End Milling process using the SLR, FL and PnP data mining algorithms. Machine feed rate is plotted along the X-axis and the Estimated Machining Time along the Y-axis. The SLR curve represents data points generated using Equation 2, which is a linear model. This model has poor estimating accuracy with an R^2 of 0.11. The negative signs in Equation 2 indicate that as the included independent variable values, i.e. D, Dc, Vc and n, increase then machining time decreases. In addition, the presence of a step function in Figure 5.1 indicates that predictions made using the SLR model significantly underestimate when actual process times are high or low. This clearly indicates the existence of non-linear relationships between the process time and independent process variable values.

The FL curve represents data points generated using Equation 4, which possesses a high estimating accuracy with an R^2 of 0.99. From Figure 5.2 and Figure 6.7 it can be seen that VM times are distributed around the FL model estimated times. This indicates the existence of a strong non-linear relationship between the independent variables and the milling machining times.

The PnP curve represents data points generated using Equation 5, which possesses a moderate estimating accuracy with an R^2 of 0.78. Figure 5.3 shows the distribution of VM times in the form of layers which indicates that the PnP method has the capability to deal with both linear and non-linear relationships. However, in order to obtain improved estimating accuracy the architecture of the neural network needs to be carefully selected. A limitation of this method is that due to the PnP method's "black box" nature it is difficult to identify the effect of individual elements of the architecture on estimating accuracy and, therefore, to select the best ANN architecture.

Figure 6. 7: Comparison of Estimating Models



The results indicate that for the Vertical End Milling process that the FL method provides a model with the highest estimating accuracy from the available data set. Further experiments were carried out in order to 'benchmark' the use of FL for developing CERs. These experiments involved validating CERs with models currently used in industry, as follows:

1. A 20% error was found when the data generated using the VM process models described in Section 4.2.2 and Section 4.2.3, were validated using a model developed by Boothroyd (2002) for the end milling process.

2. Estimating models were developed using the FL algorithm from the VM generated data as well as data generated from the use of the model. The resulting models from these experiments are shown in Table 6.8.

Table 6. 7 : Comparison of CERs Developed

	Find Laws	Simplified
CER (Boothroyd)	$T_m = (-3.83646e\text{-}006 + 1 * Lc)/Vf$	$T_m = L_c V_f^{-1}$
CER (VM process data)	$T_m = 0.998319 * Lc/(Vf - 0.00246972 * Lc)$	$T_m = (0.99L_c)(V_f - 0.002\, L_c)^{-1}$

One of the most significant achievements in the above validation process is that the Find Laws algorithm was able to derive a predictive model that is exactly equivalent to the Boothroyd's (2002) equation for estimating milling process times. Here the FL algorithm found the exact dependence of process time upon the independent process variables, and displayed this relationship explicitly in the form of a non-linear equation. It is also evident from Figure 6.7 that the FL derived models from both using data generated VM models and the Boothroyd's models produce process times that overlapped. These results validate both the VM data generation tool and the FL data analysis tool as suitable for the purpose of cost model development.

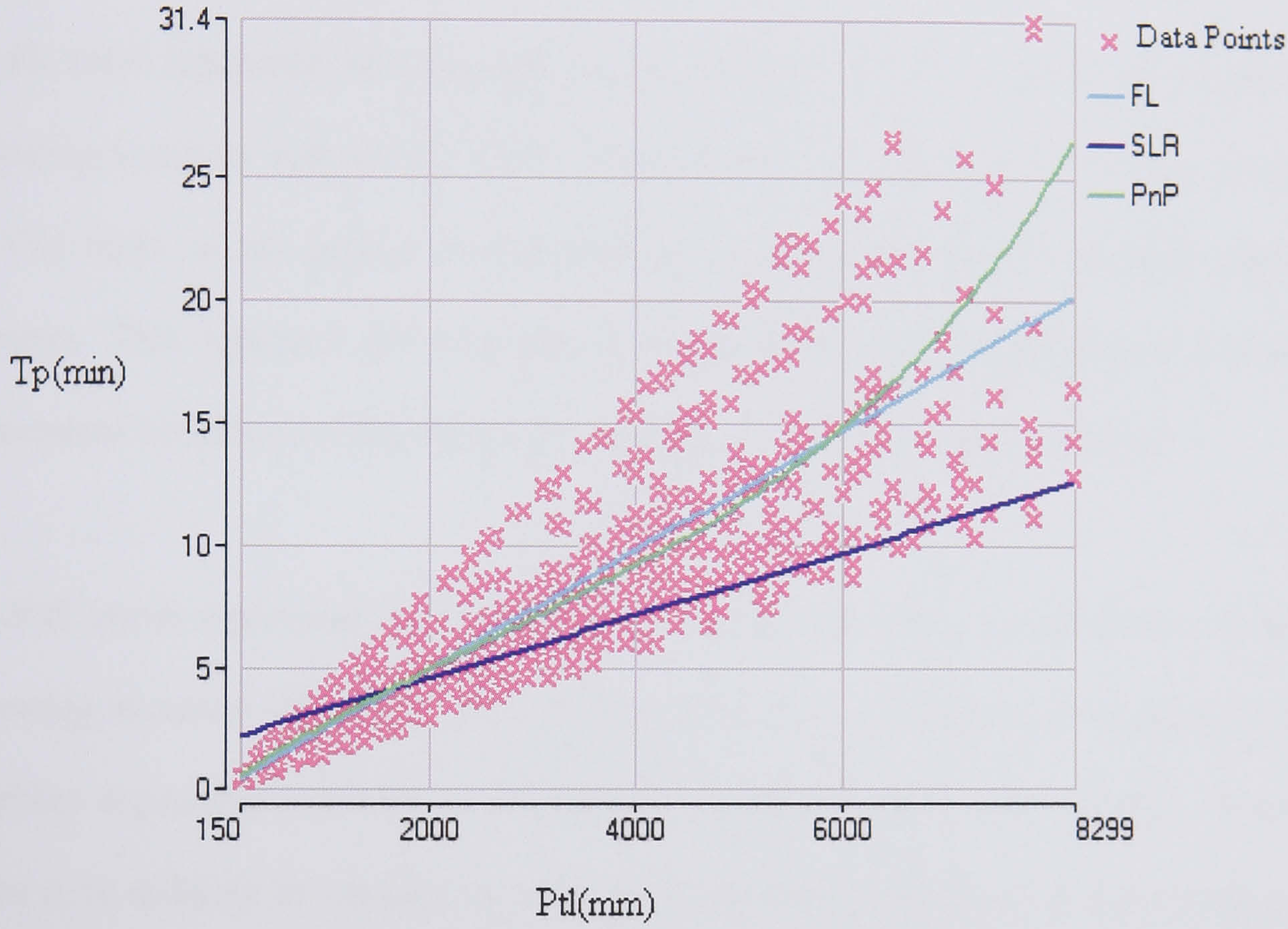
6.3.3.2 Automated Paint Spray

Figure 6.8 shows the estimating ability of the CERs developed for the automated paint spray process using the SLR, FL and PnP algorithms. The predictor variable "Path Length" was plotted along the X-axis and estimated process times along the Y-axis. The SLR curve represents data points generated using Equation 5, i.e. a linear model which, possesses a high level of estimating accuracy with an R^2 of 0.92. However, within this model the coefficients for paint fluid flow rate, paint gun speed, distance from work piece have negative effects on process times, i.e. as the values of these independent variables increase, process times decrease. In addition, the data points shown in Figure 6.4 are not distributed evenly around the predicted linear model. In particular, time estimates derived using the SLR model appear much smaller then VM times at low and high values of VM cycle time, i.e., the linear model is therefore, not valid for such cycle time values.

The FL curve represents data points generated using Equations 9 for the automated paint spray process. The estimating accuracy of this model is significantly better than that of the SLR model, i.e., with an R^2 equal to 0.98. Also, from Figure 5.7 and Figure 6.8 it can be seen that the VM times are evenly distributed around the FL derived model estimates. This indicates the existence of non-linear relationships between the independent variable values and process times. The PnP curve represents the data points generated from Equation 10. It is evident from Figure 6.8 that the PnP curve is similar to that of the FL derived model estimates again with high levels of estimating accuracy and an R^2 of 0.96. However, Equation 10 possesses approximately 24 coefficients and hence is complex making it difficult to understand the type of relationships that exist between predictor and dependent variables. Overall it can be concluded that the most appropriate process time estimating model for the Automated Spray Paint process was that developed using the Find Laws (FL) Algorithms i.e.:

$$C_{tm} = (P_{tl}(0.65+0.002G_s) + 68.17\ P_{ns}))(G_s + 1.3e\text{-}006 \times G_s^{\ 3} + 0.92\ P_{fr})^{-1}$$

Figure 6. 8: Comparison of Predictive Models for Automated Spray Paint



6.3.3.3 Turning

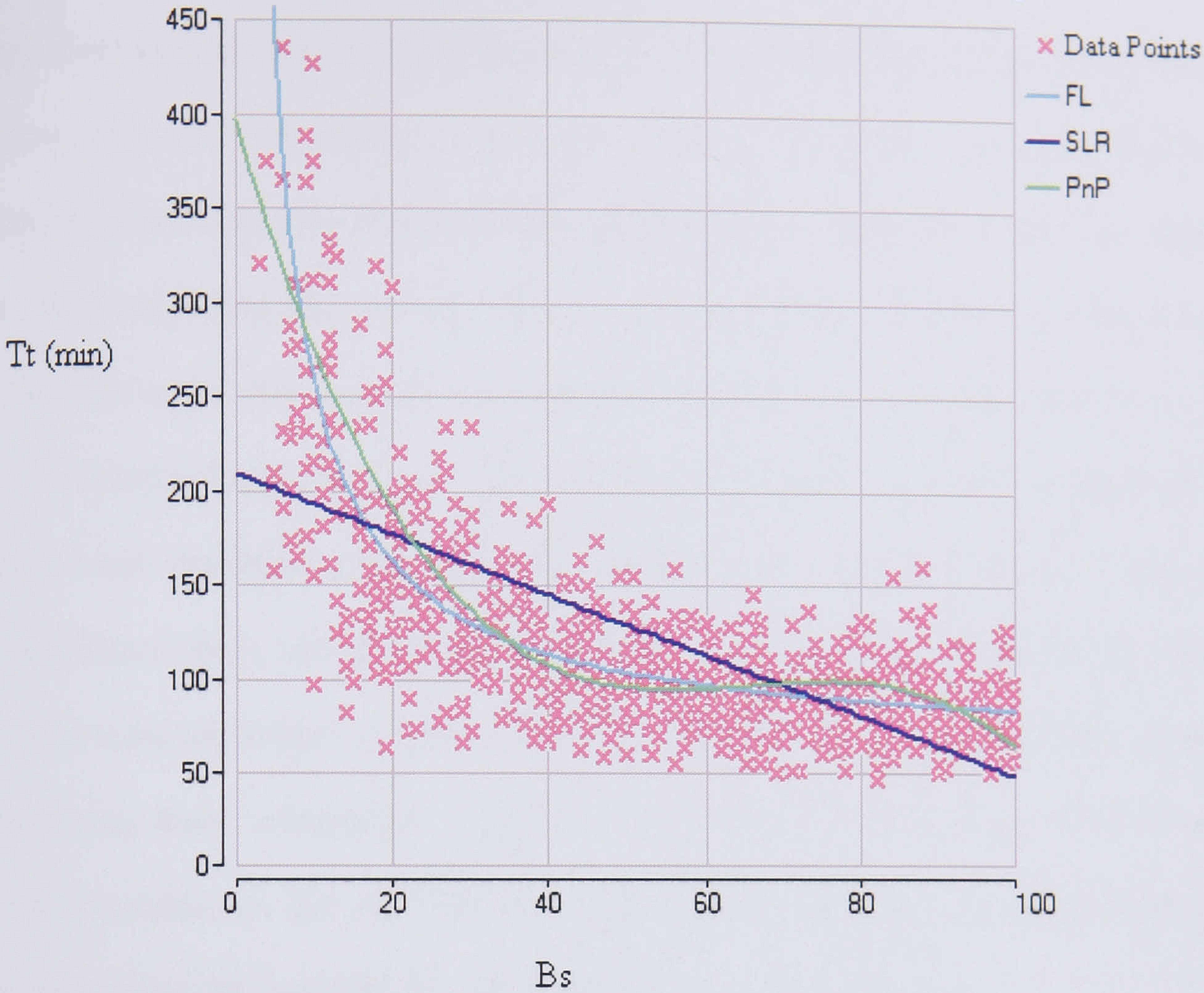
Figure 6.9 compares the CERs developed for the Turning process using the data mining algorithms of SLR, FL and PnP. "Batch size" is plotted along the X-axis and the Estimated Turning Times along the Y-axis. The SLR data is generated using Equation 11, which is a linear model. This model has poor estimating accuracy with an R^2 of 0.56. The negative coefficient for Bs in Equation 11 indicates that this variable has a positive effect in reducing machining times. It is evident from Figure 5.9 that the estimates from the SLR derived model are again much smaller than VM values at low and high values of actual process times, indicating the existence of non-linear relationships between process times and the values of the independent process variables.

The FL curve represents data generated using Equation 12, which possesses a high level of estimating accuracy with an R^2 of 0.97. From Figure 5.2 and Figure 6.9 it can be seen that the VM times in the graphs are distributed evenly around the FL derived model time estimates. This indicates the existence of a strong non-linear relationship between the independent variables and the dependent variable, i.e. milling machining time.

The PnP curve represents data generated using Equation 13, and possesses a moderate estimating accuracy with an R^2 of 0.77. The estimating relationship developed by the PnP algorithm represents a complex polynomial equation. Again, due to PnP's "black box" nature it is difficult to identify the effect of individual elements of the neural network architecture on estimating accuracy and, therefore, to select the best architecture. However, based on this observation and the MAPE values from Table 5.18, the model developed by FL appears the most suitable for estimating turning process times, i.e.:

$$\mathbf{T}=\left(293.8+\left(4.8+0.4n_t^{\,2}\right)\times t_{sb}+3.4t_{sa}+\left(3.8\,t_{ln}+13.5\,t_{pt}-5.2\right)B_s\right)\left(1+3.7B_s-0.2\,n_t\right)^{-1}$$

Figure 6. 9: Comparing Predictive Models for Turning



6.3.4 Comparing Predictive Models- II

In the previous section the predictive models developed have been examined using visual means. In Table 6.9 the models are now analysed in terms of their:

a. ability to identify cost drivers,

b. ability to identify the relationships between cost drivers and process times, and

c. accuracy of the coefficients of the cost drivers identified.

6.3.4.1 Cost Drivers Identification

A major limitations of the traditional cost model development process is its lack of methods for identifying and prioritising cost drivers. These two tasks, i.e. identification and prioritisation, are normally accomplished using subjective judgement and user expertise which often introduces bias in the resulting model estimates. Therefore, to overcome the need for such inputs this research has examined the use of the FD algorithm the result of which are listed in Tables 5.21 and 5.22. In Table 6.9 "actual" cost drivers for each of these processes were identified using the research literature (i.e. Bothroyd 2002; Wang 2000; Gopalakrishnan 1992), and discussion with process experts from Paintbox U.K. Ltd who spray paint automotive parts. It can be seen from these tables that the FD process is capable of identifying those independent process variables that have the greatest effect on the dependent variable, i.e. the cost drivers. In addition, this algorithm can also identify those data points within the complete data set that obey these dependencies.

Table 6. 8 : Comparison of DM Algorithms Predictability

	Number of Cost Drivers			Type of Relationship*			Value of Coefficient			Number of Non Cost Drivers		
	Actual	Identified	%	Actual	Identified	%	Actual	Identified	%	Actual	Identified	%
SLR												
Milling	5	4	80	5	2	40	5	3	60	3	2	66.67
Painting	4	4	100	4	4	100	5	0	0	3	3	100
Turning	16	8	50	16	8	50	16	8	50			
FL												
Milling	5	3	60	5	3	60	5	3	60	3	0	0
Painting	4	4	100	4	4	100	5	4	80	3	0	0
Turning	16	5	31	16	5	31	16	5	31			
PnP												
Milling	5	2	40	5	2	40	5	2	40	3	1	33.33
Painting	4	3	75	4	3	75	5		0	3	0	0
Turning	16	2	12.5	16	2	12.5	16	2	12.5			

*- Linear or Non –Linear Types of relationships that exist between dependent and independent variables.

The following summarises the results presented in Table 6.9, i.e.:

i) The Milling SLR model, Equation 1, has identified 4 out of the 5 actual cost drivers. However, the identified relationships between cost drivers and their coefficients identified are poor. It has also identified several non-cost drivers, the effects of which contribute to reducing levels of estimating accuracy. In terms of effect of data points, the distribution of those data points with low values dominate the whole data set and contributes to reducing estimating levels.

ii) The Spray Painting SLR model, Equation 6, identified all actual cost drivers involved in the analysis. It has also identified the correct relationships between these variables and process times. In terms of the validity of coefficients and the inclusion of non-cost drivers the SLR performance was found to be poor, i.e. with a MAPE of 30.2%.

iii) The Turning SLR model, Equation 11, identified only half of the actual variables involved in the analysis. On further observation it can be seen that all of these variables are time based and hence would have direct impact on process times. Of the identified variables there is a complete match between actual and identified relationships that exist between the dependent variable and independent variables.

iv) The FL milling predictors in Equation 3, have identified all of the cost drivers in terms of their linear/non-linear relationships between cost drivers, the validity of predictor variable coefficients, and the exclusion of non-cost driver variables.

v) The Painting FL model, Equation 8, identified all relevant cost drivers, their linear/non-linear causal relationships and their coefficients with greater precision than the SLR process, i.e. with a MAPE of 5.1%.

vi) The Turning FL model, Equation 12, has identified approximately a third of the actual cost drivers, i.e. 5 from 16, that have direct impact on process times. The accuracy of this prediction is evident from Table 6.6, which shows that the FL process is consistent in its overall performance.

vii) The estimating accuracy of PnP derived models seem surprisingly poor, since PnP has the ability to deal with both linear and non-linear relationships, and should, therefore, in theory be able to provide better estimating accuracy. However, this was not found to be the case. These unexpected results from PnP could have arisen, because the level of estimating accuracy is dependent on selection of appropriate neural network architectures. In this particular case, PnP chooses it architecture automatically, i.e. there is no provision for it to be modified.

6.4 Summary

6.4.1 Application of CMD Process in Industry

Results obtained indicate that this methodology can be applied over a wide range of manufacturing processes such as machining, painting, fabrication, welding, for both batch and process manufacturing and at different levels of process and product detail (i.e. Table 3.5 and 3.6), where,

1) There is few/ no historical data available,
2) Level of expertise required is low, i.e. a high level of expertise in the areas of data identification, data collection and data analysis are not necessarily required since a

person who only has knowledge and understanding of manufacturing processes and virtual process models could develop CERs.,

3) Little information about the manufacturing process and cost drivers is available,
4) Process in its conceptual stage,
5) Also it can be said that with these little information and less expertise, the results obtained were so good that these method can be extended to be used at detailed design and operational stage of the process.

6.4.2 Advantages

1. It is a one-off model building process, modifications can be made to suit the requirement to the existing models and hence estimate the models.
2. It is not time consuming when used for more than one task. For example, modifying an existing VM model to enable addition process trials and/or design features to be costed may only take relatively short periods, (i.e. days) when compared with original model development time, (i.e. weeks)
3. Reduced reliance on expertise.

6.4.3 Disadvantages

Learning time of virtual manufacturing simulation software was found to be the only time consuming task in the whole process of CMD, i.e. two to four weeks is normally required for training on the complete VM software package.

Where a new process or product is being developed there is little historical data available unless the process or product is a modified version on an existing one. Traditional data collection methods rely on data being available and merely enable the data source to be assessed and the relevant data extracted from their source. All such methods rely on data being available that links process activities, process features and product features with their associated cost and/or process times.

Chapter 7 Conclusions

This research has examined the need for new methods for developing cost models for products and processes at the conceptual stage of their development. Here the work has involved investigating the use of virtual manufacturing and data mining to help resolve the need for high levels of subjective expertise and lack of data from which to develop valid CERs. The main conclusions of this work are as follows:

1. Virtual Manufacturing has been identified as a suitable method for generating data from which valid cost models may be developed. Here experiments were undertaken that both validated the virtual manufacturing process models themselves and the data generated from them.

2. The Taguchi Methodology that makes use of Orthogonal Arrays has been identified as a suitable method for designing Virtual Manufacturing data generation trials that ensure the data generated is fit for purpose.

3. The data mining process has been identified as an advanced data analysis tool, which meets the requirements of the existing, cost model development process. Experiments have been carried out in order to demonstrate its ability to build predictive Linear and Non-linear CERs. Experiments have also been carried out in

order to demonstrate the ability of data mining to identify the cost drivers of a particular process.

4. Although data mining has been identified as being capable of generating CERs that provide accurate estimates, the results of this work have brought into question the validity of CERs and hence the ability of data mining process as an automatic data analysis technique. A major factor affecting the capability of the data mining process is the levels of process complexity that exist in terms of the number of variables that influence costs.

5. Experiments were undertaken in order to identify the effects of Number of Variables and Number of Data Points on the estimating accuracy of the cost models derived using data mining techniques. It has been found that with increasing Numbers of Data Points and Variables accuracies of the FL and PnP algorithms has increased. However, SLR tends to exhibit a mixed behaviour in this case when non-linear relationships exist amongst the predictor variables.

6. Cost models can be generated automatically using virtual manufacturing for data generation followed by data mining to develop CERs. This has been achieved for a simple mature technology of milling and the complex process of automated paint spraying where the process technology is continuing to develop.

Chapter 8 Further Work

The following further work areas have been identified from the research undertaken:

1- Further work needs to be undertaken in order to test the suitability of the developed methodology over a wider spectrum of manufacturing process applications.

2- The work undertaken within this research has made significant contributions to the cost model development process in terms of data generation and to overcome the problem of lack of data. However, further work is needed to establish the validity of data generated for those new manufacturing processes where benchmarking tools are not available for assessment purposes. In addition, the development of Virtual Manufacturing process models was found to be a time consuming task within the process of cost model development. This needs to be explored further in order to identify method through which Virtual Manufacturing process models can be developed in shorter time periods. Although VM has been found to be suitable for generating data at different levels of process detail, it is difficult to model process features such as the temperature in a paint booth. Hence, further work needs to be carried out in order to find ways of including these types of process features in the VM model.

3- In addition, the development of Virtual Manufacturing process models was found to be a time consuming task within the process of cost model development. This needs to be explored further in order to identify method through which Virtual Manufacturing process models can be developed in shorter time periods. However, with the use of this methodology time to develop cost models can be greatly reduced, for example if time taken to develop cost models for a particular process using traditional CMD methods is four weeks the same cost models could be developed in two-three weeks time using this methodology i.e. almost 50 to 25% reduction in cost model development time can be achieved.

4- Although the data mining process has been found to be capable of developing valid CERs further work needs to be carried out to develop CERs when there are categorical variables such as product material, involved in the process.

5- This methodology of cost model development is rationally helpful with developing process technologies. This is the case because as manufacturing process technology develops the existing data is no longer valid and there are fewer comparatives to benchmark the results against industry standards with leading edge technology. Experts are also more difficult to find – if they exist at all. Further work, therefore, needs to be undertaken to establish ways of validating the CERs developed for the newer manufacturing processes and materials.

6- A possible exploitation opportunity for this cost model development process could be as a validation tool, as the product develops, for appraising quotations for out-sourced components where detailed knowledge of the supply chain technology is absent. It could be possible for a team of process experts in a particular process technology field to provide a service for quotation appraisal using the VM/DM techniques. In this instance further work to produce a fully integrated user friendly automated cost modelling system would be required.

References

Acosta, C, Switek, W., and Garcia Y.E., 1997, Dynamic modeling in turning machining, *Proceedings of the 21st international conference on Computers and industrial engineering*, San Juan, Puerto Rico, pp.397 - 400.

Alabastro, M.S., Beckmann, G., Gifford, G. et al, 1995, The use of visual modeling in designing a manufacturing process for advanced composites structures. *IEEE Transactions On Engineering Management*, Vol.42, No.3, pp.233-242.

Anna, B., 1995, Work Measurement- the Basic principles revisited, *Work Study*, Vol.44, No.7, pp.10-17.

Antony, J., Somasundarum, V., Fergusson, C., Blecharz, P., 2004, Applications of Taguchi approach to statistical design of experiments in Czech Republican Industries, *International Journal of Productivity and Performance Management*, Vol.53, No.5/6, pp.447-457.

Apgar, H., 2003, Developing the cost estimating model, *AIAA Space 2003 Conference & Exposition*, 23-25 Sept 03, Vol.2, AIAA-2003-6343, Long Beach, California.

Asakawa, N., Takeuchi, Y., 1997, Teachingless spray-painting of sculptured surface by an industrial robot, *Proceedings of the 1997 IEEE, International conference on robotics and automation*, Albuquerque, New Mexico, pp.1875-1879.

Baguley, P., 2004, Improving the cost model development process using fuzzy logic. Thesis (Ph.D) - De Montfort University, Leicester, U.K.

Baguley, P., Shaik, T., Fresco, J. and Stockton, D.J., 2003a, Cost modeling with fuzzy logic, *20th International Manufacturing Conference: Knowledge Driven Manufacturing*, Cork Institute of Technology, Cork, Ireland.

Baguley, P., Wang, Q. and Stockton, D.J., 2000, Applied advanced cost estimating techniques for manufacturing processes, *Proceedings of the 17th Conference of the Irish Manufacturing Committee*, Galway, Ireland, August 2000, pp.335-340.

Baines, T.S., Harrison, D.K., 1999, An opportunity for system dynamics in manufacturing system modelling, *Production Planning & Control*,Vol.10, No.6, pp.542-552.

Balakin, K.V. and Savchuk, N.P., 2004, Advanced data mining technologies, *Manufacturing Chemist*, Vol.75, No.7, pp.31-33.

Banks, J., 1999, Introduction to simulation, in: e.a. P.A. Farrington (Editor), *Proceedings of the 1999 Winter Simulation Conference*, Atlanta, U.S.A, pp. 7-13.

Banks, J., 2000, Simulation in the future, panel discussion, Proceedings of the Winter Simulation Conference, pp. 1568-1576.

Barnes, R. M., (1980), Motion and Time Study Design and Measurement of Work, John Wiley & Sons, Chichester, pp. 376.

Barrett, D.S., 1997, Standard data in batch production at Rank Xerox. *Work Study*, Vol.46, No.1, pp.9-19.

Bode, J., Ren, S. and Shi, Z., 1995, Application of 3-Layer perceptrons to cost estimation, *Proceedings of the 1995 IEEE International Conference on Neural Networks*, pp.1749-1762, ISSN: 0-7803-2768-3.

Bode, J., 2000, Neural networks for cost estimation: simulations and pilot application, *International Journal of Production Research*, Vol.38, No.6, pp.1231-1254.

Boothroyd, G., Dewhurst, P. and Knight, W., 2002, Product design for manufacture and assembly, Marcel Dekker, New York, ISBN: 0-8247-0584-X, pp.277-282,313.

Brown, J., 2004, Digital Manufacturing: The PLM approach to better Manufacturing processes. Tech-Clarity, Inc.

Buchner, A.G., Anand, S.S., & Hughes, J.G., 1997, Data Mining in Manufacturing Environments: Goals, Techniques and Applications, *Studies in Informatics and Control*, Vol.6, No.4, ISSN 1220-1766, pp. 319-328.

Bulkley, J., Gayle, S., Hicks, B. and Stephens, R., 1999, Adding the where to the who, *Proceedings of the 24th Annual SAS Users Group International Conference*, Miami, Florida, USA, pp. 1-4.

CIMdata, 2003, The Benefits of Digital Manufacturing, CIMdata Inc., Michigan.

Clark, D.O., 1977, The MTM Family of systems, *Society of Manufacturing Engineers- the Computer and automated systems association*, MS77-325, pp.15.

Cochran, E.B., 1976, Using regression techniques in cost analysis. Part 2, *International Journal of Production Research*, Vol.14, No.4, pp.489-511.

Cohen, Y., Bidanda, B. and Billo, R.E, 1998, Accelerating the generation of work measurement standards through automatic speech recognition: a laboratory study, *International Journal of Production Research*, Vol.36, No.10, pp.2701-2715.

Creese, R.C., Adithan, M. and Pabla, B.S., 1992, Estimating and costing for the metal manufacturing industries, Marcel Dekker, Inc, New York, pp.120.

Curran, R., Kundu, A., Crosby, S. and Raghunathan, S., 2003, Concept modelling for aircraft manufacturing, *2003 AACE International Transactions*, pp. EST.21.1-12.

Dauda, M., Roy, R. and Stephenson, D.J., 2005, Development of a cost model for high efficiency deep grinding (HEDG), In: J.X. Gao, Baxter, D., Sackett, P. (Editor), ICMR2005: *3rd international conference on manufacturing research incorporating the 21st national NCMR*, Cranfield University, U.K.

Delgado, Y., McNeill, M and Stockton, D.J., 2002, Improving the cost model development procedure, *2002 AACE International Transactions*, Chicago, pp. EST.14.

Denton, J., and M. Hung, 1996, A comparison of nonlinear optimization methods for supervised learning in multilayer feedforward neural networks, *European Journal of Operational Research*, Vol. 93, pp 358-368.Creese, R.C., 1999, Introduction to Manufacturing Processes and Materials, Marcel Dekker, New York, ISBN: 0-8247-9914-3.

Dysert, L.R., 1999, Developing a parametric model for estimating process control costs, *AACE International transactions*, ABI/INFORM Global, Morgantown, pp. Est.01.1-4.

Endo, J., Ohba, S. and Anzai, T., 1996, Virtual manufacturing for sheet metal processing, *Journal Of Materials Processing Technology*, Vol.60, Pp.191-196.

Fazel, F.A., Ouyang, J., 2003, Data mining: Understanding data and disease modeling, *IASTED International Multi-Conference on Applied Informatics*, Vol.21, pp.32-37.

Grewal, S. and Choi, C.K., 2005, An integrated approach to manufacturing process design and costing, *Concurrent Engineering: Research & Application*, Vol.3, No.3, pp.199-207.

Gibbs, D., 1984, An introduction to CNC machining, Cassell Computing, Paston Press, Norwich, ISBN: 0-304-31169-3.

Giudici, P., 2003, Applied data mining: Statistical methods for business and industry, John Wiley, ISBN: 0470846798.

Gorlani, C. and Zavanella, L., 1993, Continuous simulation and industrial processes: electrode consumption in arc furnaces, *International Journal of Production Research*, Vo1.31, No.8, pp. 1873-1890, ISSN 0020-7543/93.

Graham, P., 2004, Fuzzy logic and neural nets: still viable after all these years?, EDN, Vol.49, No.12, pp.69-72.

Haigang, L., 2005, Applications of data warehousing and data mining in the retail industry, *2005 International Conference on Services Systems and Services Management (IEEE Cat. No. 05EX1010), 13-15 June 2005. IEEE*, Piscataway, NJ, USA, pp. 1047-50 Vol. 2.

Hand, D., Mannila, H. and Smyth, P., 2001, Principles of Data Mining, The MIT Press, London, ISBN: 0-262-08290-X.

Holst, L., 2001, Integrating Discrete-Event Simulation Into The Manufacturing System Development Process, Lund University, Sweden, ISBN 91-631-1404-6. Hollocks, B. W., 1995, The Impact of Simulation in Manufacturing Decision Making, *Control Engineering Practice*, Vol. 3, No. 1, pp. 105-112, Elsevier, Great Britain.

Hosking, J.R.M, Pednault, E.P.D. and Sudan, M., 1997, A statistical perspective on data mining, *Future Generation Computing Systems*, special issue on Data Mining, 1997.

Heilala, J., 1999, Use of simulation in manufacturing and logistics systems planning, AS116.140, *VTT Manufacturing Technology*, Finland.

http://www.bsonline.bsi-global.com/server/index.jsp, [Access on September, 2004].

http://www.delmia.com/, 2004. Delmia D5R13 VNC and UPaint, Delmia.

http://www.fraisa.com/, 2004. Roughing End Mills- Cutting Data.

Hu, P.-H., and Ehmann, KF, 2000, A Dynamic Model of the Rolling Process - I. - Homogeneous Model, *International Journal of Machine Tools and Manufacture*, Vol. 40, No.1, pp. 1-19

Hundal, M.S., 1997, Product costing: A comparison of conventional and Activity-based costing methods, *Journal of Engineering Design*, Vol.8, No.1, pp.91-104.

Idri, A., Khoshgoftaar, T., Abran, A. and Robert, S., 2002, Can Neural Networks be easily Interpreted in Software Cost Estimation?, *Fuzzy-IEEE*, 12-17 May, Hawai, 2002, pp. 1162-1166.

Ince, R., Hutton, R. and Whitehead, R., 2001, Rapid Cost modelling for the small and medium sized company, *Management services*, pp.1-10.

Jahan-Shahi, H., Shayan, E. and Masood, S., 1999, Cost estimation in flat plate processing using fuzzy sets, *Computers & Industrial Engineering*, Vol.37, pp.485-488.

Jayaraman, A. and Gunal, A.K., 1997, Application of Descrete Event Simulation in the Design of Automotive Powertrain Manufacturing Systems, In Andradóttir, S., Healy, K.J., Withers, D.H., and Nelson, B.L. (Eds), Proceedings of the 1997 Winter Simulation Conference. 7-10 December 1997, Atlanta, GA, pp. 758-764.

Jayaram, S., Connacher, H.I. and Lyons, K.W., 1997, Virtual assembly using virtual reality techniques, Computer-Aided Design, Vol.29, No.8, pp 575-584.

Jung, J.Y., 2002, Manufacturing cost estimation for machined parts based on manufacturing features, *Journal of Intelligent Manufacturing*, Vol.13, pp.227-238.

Kapoor, L.M., 1990, Determining life cycle costs of a work measurement system (WMS). *Aerospace and Electronics Conference, NAECON 1990, Proceedings of the IEEE 1990 National*, pp. 995-1000.

Kalpakjian, S., 1995, Manufacturing Engineering and Technology, 3rd Edition, Addison-Wesley publishing company, pp.1231-1236, ISBN :0-201-84552-0.

Kasravi, K., 1997, Data Mining and Knowledge Discovery in Manufacturing, AUTOFACT 97, Nov 3-6, 1997, *Published by Society of manufacturing Engineers*, Dearborn, MI., Detroit, Michigan.

Kay, R., 2004, Fuzzy logic. (Quick Study), *Computer world*, Vol.38, No. 35, pp.26 (1).

Kibira, D., McLean, C., 2002, Virtual reality simulation of a mechanical assembly production line, In: C.H.C. E.Yucesan, J.L. Snowdon, and J.M. Charnes (Editor), proceedings of the 2002 Winter simulation conference.

Kilgore, J.T., 1997, Standard data: developing an effective predetermined time system, *IIE Solutions*, Vol.29, No.6, pp.40-42.

Kim, G., An, S. and Kang, K., 2004, Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning, *Building and Environment*, Vol. 39, pp. 1235-1242.

Kim, J., Dronfeld, D.A., 2001, Cost estimation of drilling operations by a drilling burr control chart and bayesian statistics, *Journal of Manufacturing Systems*, Vol.20, No.2, pp.89-97.

Klein, R., 2002, A new tool for improving spray gun transfer efficiency, *Metal Finishing*, August 2002, pp.34-41.

Klein, G.A., 1993, A recognition-primed decision (RPD) model of rapid decision making, in G.A. Klein et al. (Eds), Decision Making in Action: Models and Methods, Ablex, N.J., pp.138-137.

Knooce, D., Judd, R., Sormaz, D., et al., 2003, A hierarchical cost estimation tool, *Computers in Industry*, Vol.50, pp.293-302.

Knobbe, A., 1997, Application of data mining, *Informatie*, Vol.39, pp.7-10.

Krar, S.F., Rapisarda, M. and Check, A.F., 1998, Machine Tool and Manufacturing Technology, Delmar Publishers, London, ISNB: 0-8273-6351-6.

Kusiak, A., 1995, Engineering Design; Products, Processes, and Systems, Academic Press, London, ISBN: 0-12-430145-2.

Law, A.M., and Kelton, W.D., 1991, Simulation modelling and analysis, 2nd ed. McGraw-Hill, Chichester.

Law, A.M. and McComas, M.G., 1998, Simulation of manufacturing systems, *Proceedings of the 1998 Winter Simulation Conference*, Averill M.Law & Associates Inc., Arizona, U.S.A.

Lawrence, P.J., 2003, The Multiple roles of discrete event simulation in the workplace, *Global Journal of Engineering Education*, Vol.7, No.2, pp.165-172.

Lee, J.D. and Moray, N., 1994, Trust, Self-confidence, and operators' adaptation to automation, *International Journal of Human-Computer Studies*, Vol.40, pp.153-184.

Lee,W.B., Cheung, C.F. and Li, J.G., 2001, Applications of virtual manufacturing in materials processing, Journal of Materials Processing Technology, Vol.113, pp.416-423.

Levine, H. L. and Latham, P., 2004, The use of models to estimate and control the cost of biopharmaceutical manufacturing, *BioProcess Technology Consultants,* Inc.

Lin, E., Minis, I., Nau, D.S. and Regli, W.C., 1996, An Assessment of Virtual Manufacturing Technologies, *Proceedings of the 1996 NSF Design and Manufacturing Grantees Conference,* University of New Mexico, Sandia National Laboratories, Albuquerque, New Mexico, Jan. 2-5, 1996, pp. 101-102.

Ling, C.X. and Chenghui, Li., 1998, Data mining for direct marketing: problems and solutions, *Proceedings Fourth International Conference on Knowledge Discovery and Data Mining*, 27-31 Aug. 1998. AAAI Press, New York, NY, USA.

Luxhoj, J.T. and Giacomelli, G.A., 1989, Comparison of labor standard for a green house tomato production system; A Case Study, *International Journal Of Production Management*, Vol.10, No.3, pp.38-49.

Mackenzie, D., 2003, Cost estimating relationship regression variance study, *AIAA Space 2003 Conference & Exposition*, 23-25 Sept 03, Vol.2, AIAA-2003-6297, Long Beach, California.

Madhisetty, S.R., 2004, Data Mining and its applications, *2004 IRMA International Conference*, pp. 1120-1122.

Martin, J.C., 1975, How 4M data improves assembly methods and standards, *Society of Manufacturing Engineers*, AD75-774, pp.1-15.

Mason, A.K. and Khan, D. J., 1997, Estimating costs with fuzzy logic, *AACE International Transactions 1997*, EST.03.1-6.

Maynard, (1948) in Zandin, K.B., 2002, Most Work Measurement Systems. Marcel Dekker, 1-7 pp.

Matko, D., Zupancic, B. and Karba, R., 1992, Simulation and modelling of continuous systems; A case study approach, Prentice Hall, ISBN: 0-13-808064-X.

Melin, J.B., 1994, Parametric Estimation, Cost Engineering, Vol. 36, No.1, pp.19-23.

Mileham, A.R., Currie, G.C., Miles, A.W., and Bradford, D.T., 1993, A parametric approach to cost estimating at conceptual stage of design, *Journal of Engineering Design*, Vol.4, No.2, pp.117-125.

MINITAB Release 14, S.S., 2004.

Mitra, S., Pal, S.K. and Mitra, P., 2002, Data mining in soft computing Communicating framework survey, *IEEE Transactions on Neural Networks*, Vol. 13, No.1, pp.3-14.

Mujber, T.S., Szecsi, T., and Hashmi, M.S.J., 2004, Virtual reality applications in manufacturing process simulation, *Journal of Materials Processing Technology*, Vol.155-156, Pp.1834-1838.

Nakayama, S., Nakayama, K. and Nakayama, H., 2002, A study on setting standard time using work achievement quotient, *International Journal of Production Research*, Vol.40, No.15, pp.3945-3953.

Needy, K.L., Billo, R.E., and Warner, R.C, 1998, A cost model for the evaluation of alternative cellular manufacturing configurations, *Computers Industrial Engineering*, 34(1): 119-134.

Nouldus, 2004, Measuring behavior in virtual environment, Nouldus Information Technology, www.Nouldus.com, [Access on August 2004].

Oscarsson, J., 2000, Enhanced Virtual Manufacturing, De Montfort University, Leicester, pp.120.

Perera, T., and Liyanage, K., 2001, IDEF based methodology for rapid data collection, *Integrated Manufacturing Systems*, Vol.12, No.3, pp.187-194.

Petri, L.K., Billo, R.E. and Bidanda, B., 1998, A neural network process model for abrasive flow machining operations, *Journal of Manufacturing Systems,* Vol.17, No.1, pp.52-64.

Ping, L., Yongtong, H., Bode, J., Shouju, R., 1996, Multi-agent system for cost estimation, *Computers & Industrial Engineering*, Vol.31, No.3/4, pp.731-735.

Popham, K., 1996, Cost estimating using historical costs, *AACE International Transactions*, Morgantown, Est.14.1-5.

Randal, S.Z., 1996, Accurate estimate start with clear pictures, Transactions of AACE International. CS101.

Rocke, D., 1998, Perspective on Statistical Tools for Data Mining Applications (In Proceedings), *Proceedings of the Second International Conference on Practical Application of Knowledge Discovery and Data Mining*, pp. 313-318.

Rohrer, M.W., 2000, Seeing is Believing: The importance of visualization in manufacturing simulation., *Proceedings of the 2000 Winter Simulation Conference*, Bountiful, UT, U.S.A., pp. 1211-1216.

Roy, R., Forsberg, S., Kelvesjo, S., and Rush, C., 2001, Quantitative and qualitative cost estimating for engineering design, *Journal of Engineering Design*, Vol.12, No.2, pp.147-162.

Roy, R., Forsberg, S., Kelvesjo, S., and Rush, C., 2000, Risk analysis of parametric cost estimates within a concurrent engineering environment, 7th ISPE *International Conference on Concurrent Engineering: Research and Applications*, Lyon, France, July 17th - 20th, Technomic Inc., Pennsylvania USA, pp.23-30.

Rowlands, H., Antony, J., Knoles, G., An application of experimental design for process optimisation, *The TQM Magazine*, Vol.12, No.2, pp.78-83. Read, B.J., 1999, Data mining and science? , 12th ERCIM Workshop on Database Research, Amsterdam, 2-3 Nov 1999, pp.1.

Rush, C., and Roy, R., 2001a, capturing quantitative & qualitative knowledge for cost modelling within a CE environment, *ISPE International Conference on Concurrent Engineering: Research and Applications*, Anaheim, Los Angeles, July 28th - August 1st, CETEAM, Pennsylvania USA, pp. 209-218.

Rush, C., Roy, R., 2001b, Expert judgment in cost estimating: Modelling the reasoning process. *Concurrent Engineering: Research and Applications* (CERA) *Journal*, Vol.9, No.4, December 2001.

Rudas, I. J. and Horvath, l. (1997), Modeling of manufacturing processes using a Petri net representation, *Engineering Applications of Artificial Intelligence,* vol. 10, No. 3, pp. 243-255.

Sadoyan, H., Zakarian, A., Mohanty, P., 2006, Data mining algorithm for manufacturing process control, *International Journal of Advanced Manufacturing Technology*, Vol.28, No.3-4, pp.342-350.

Salas, A., 2004, Cost Estimation: More than an analytical tool, *2004 IEEE Aerospace Conference Proceedings*, pp.1053-1060.

Scalan, J., Hill, T., et al, 2002, Cost modelling for aircraft design optimization, *Journal of Engineering Design*, Vol.13, No.3, pp.261-269.

Schreve, K., Schuster, H.R., and Basson, A.H., 1999, Manufacturing cost estimation during design of fabricated parts, *Proceedings of Institution of Mechanical Engineers*: IMechE, 213(part B), pp.731-735.

Seifert, J. W., 2004, Data Mining: An Overview, In CRS Report for Congress, http://www.fas.org/irp/crs/RL31798.pdf, [Access on March 2005].

Setyawati, B.R., Creese, R.C., and Shirman, S., 2003, Neural networks for cost estimation (part2), *2003 AACE International Transactions*, pp. Est.14.1-10.

Shatla, M. and Altan, T., 2000, Analytical modeling of drilling and ball end milling, *Journal of Materials processing Technology*, January (2000), pp.125-133.

Shehab, E.M., Abdalla, H.S., 2001, Manufacturing cost modelling for concurrent product development, *Robotics and Computer-Integrated Manufacturing*, Vol.17, No.4, pp.341-353.

Shehab, E.M. and Abdalla, H.S., 2002, A design to cost system for innovative product development, Part B: *Journal of Engineering Manufacture*, pp.999-1018.

Shin, J.G., Lee, K.K., Woo, J.H. and Kim, W.D. et al, 2000, Modelling and simulation of production process in subassembly lines at a shipyard, *Research Institute of Marine Systems Engineering*, Seoul, Korea.

Shtub, A., Zimerman, Y., 1993, Neural-network-based approach for estimating the cost of assembly systems, *International Journal of Production Research*, Vol.32, No.2, pp.189-207.

Shtub, A. and Versano, R, 1999, Estimating the cost of steel pipe bending, a comparison between neural networks and regression analysis, *International Journal of Production Economics*, Vol.62, pp.201-207.

Smith, A.E., Mason, A. K., 1997, Cost estimation predictive modeling: Regression versus neural network, *The Engineering Economist*, Vol.42, No.2, pp.137-161.

Son, Y.K., 1991, A cost estimation model for advanced manufacturing systems, International Journal of Production Research, Vol.29, No.3, pp.441-452.

Song, J.-S., Lee, T.-E., 1998, Petri net modeling and scheduling for cyclic job shops with blocking, In: *Computers and Industrial Engineering,* Volume 34, Issue 2, pages 281-295.

Sonmez, R., 2004, Conceptual cost estimation of building projects with regression analysis and neural networks, *Canadian Journal of Civil Engineering*, Vol.31, pp.677-683.

Stewart, R.D., Wyskida, R.M., and Johannes, J.D., 1995, Cost Estimator's Reference Manual, 2nd, John Wiley & Sons, Inc. Chichester.

Stockton, D.J., Delgado, Y, McNeill, M, and Baguley, P, 2001, Improving the cost model development process (2), De Montfort University, EPSRC Ref. No.: GR/M 58818-2, Leicestershire.

Stockton, D.J., Delgado, Y, McNeill, M, and Baguley, P, 2000, Improving the cost model development process (1), De Montfort University, EPSRC Ref. No.: GR/M 58818-1, Leicestershire.

Stockton, D.J. and Middle, J.E., 1982, An approach to improving cost estimating, *International Journal of Production Research*, Vol.20, No.6, pp.741-751.

Stockton, D.J. and Wang, Q., 1999, Applying advanced modelling techniques to cost estimating, *Conference of Association of Cost Engineers: Engineering Manufacturing Committee*, De Montfort University, Leicester,U.K.

Switek, W. and Majewski, T., 1997, Dynamic modeling and optimization for technology management, *Proceedings of the 21st international conference on Computers and industrial engineering*, San Juan, Puerto Rico, pp.11-14.

Szekely, J. and Trapaga, G., 1994, Some Perspectives on the Mathematical Modelling of Materials Processing Operations, *Modelling and Simulation in Materials Science and Engineering*, Vol. 2, pp. 809-828.

Thamaraiselvi, G. and Kaliammal, A., 2004, Data Mining: Concepts and Techniques, *SRELS Journal of Information Management*, Vol.41, No.4, pp.339-348.

Todorovski, L., Bridewell, W., Shiran, O. and Langley, 2005, Inducing Hierarchical process models in dynamic domains, American Association for Artificial Intelligence, pp.1-6.

Ulgen, O. and Gunal, A., 1998, Simulation in the Automobile Industry, John Wiley and Sons, Inc., ISBN 0-471-13403-1, pp.547.

Ulgen, O. and Gunal, A., 1996, Pitfalls of simulation modelling and how to avoid them by using a robust simulation methodology, AutoSimulations Symposium 96 Proceedings.

Unal, R. and Dean, E.B., 1991, Taguchi approach to design optimization for quality and cost: An overview, In *1991 Annual Conference of the International Society of Parametric Analysis*, pp.1-10.

Verlinden, B., Collin, P., Duflou, J.R. and Cattrysse, D., 2005, Cost estimation for sheet metal parts: A case study in: J.X. Gao, Baxter, D., Sackett, P. (Editor). ICMR2005: 3rd *International conference on manufacturing research Incorporating the 21st National NCMR*, Cranfield University.

Wang, Q., 2000, Improving the cost model development process using neural networks, PhD Thesis, De Montfort University, Leicester, pp.179p.

Wang, Q. and Stockton, D.J., 2001, Cost model development using artificial neural networks, Aircraft Engineering & Aerospace Technology, Vol.73, No.6, pp.536-541.

Wang, Q. and Stockton, D.J., 2003, Artificial Neural Networks for Improving cost model development process, *International Journal for Manufacturing Science and Production*, Vol.5, No.1-2, pp.91-94.

Wei, Y. and Egbelu, P.J., 2000, A framework for estimating manufacturing cost from geometric design data, *International Journal of Computer Integrated Manufacturing*, Vol.13, No.1, pp.50-63.

Weustink, I.F., Brinke, E.T., Streppel, A.H. and Kals, H.J.J., 2000, A generic framework for cost estimation an cost control in product design, *Journal of Materials Processing Technology*, Vol.102, pp.141-148.

Whiteside II, J.D., 2003, Cost model performance, *AACE International Transactions*, Morgantown, pp. EST.11.1-11.

Williams, E.J., 1997, Discrete-Event Process Simulation for the Continuous Simulation Model Builder. In *Proceedings of the 1997 Summer Computer Simulation Conference*, (Eds) Mohammad, S. O. and John I., pp. 362-367.

Wood, R.L., 1999, Recognising how humans make process modeling difficult, *The Institution of Electrical Engineers*, Vol.7, pp.1-5.

Worn, I.H., Frey, D. and Keitel, J., 2000, Digital Factory- Planning and running enterprises of the future, Proceedings of the 26th Annual Conference of the IEEE Electronics Society IECON-2000, pp. 1286-1291., ISSN 0-7803-6456-2.

Zhang, Y.F., Fuh, J. Y. H. and Chan W. T., 1996, Feature-based cost estimation for packaging products using neural networks, *Computers in Industry*, Vol.32, No.1, pp.95-113.

Zhao, W., Li, J., Zhang, C., Liang, Z., and Wang, B., 2006, Feature-based cost estimation for composite structure with fuzzy Multi-Attribute Utility Theory, *International Journal of Industrial and Systems Engineering*, Vol.1, No.1/2, pp.284-300.

Bibliography

1) Ahluwalia, R.C., S., 2004, Selection of Data analysis techniques for data mining applications, *Proceedings of SPIE,* Vol.5605, Bellingham, pp. 97-108.

2) Askin, R.G. and Standridge, C.R., 1993b, Modeling and Analysis of Manufacturing Systems. John Wiley & Sons, Inc, Chichester, pp.17-27.

3) Baguley, P., Maropoulos, P. and Page, T., 2004, Cost engineering for new product design and innovation, *Society for Manufacturing Engineers*, TP04PUB302: 6.

4) Banerjee, P., Kesavadas, T., et al, 1999, Industrial Virtual Reality, MED-Vol.9, ASME, New York.

5) Belmiro, T.R. and Pina, A.A.D., 2001, A Process Modeling Approach at Xerox of Brazil. *Work Study*, Vol.50, No.7, pp.269-275.

6) Ben-Arieh, D., 2000a, Cost estimation system for machined parts, *International Journal of Production Research*, Vol.38, No.17, pp.4481-4494.

7) Ben-Arieh, D., Lavelle, J.P., 2000b, Manufacturing Cost Estimation: Applications and Methods, *Engineering Valuation and Cost Analysis*, Vol.3, pp.43-55.

8) Ben-Arieh, D., Li, Q., 2003, Web-based cost estimation of machining rotational parts, *Production Planning & Control*, Vol.14, No.8, pp.778-788.

9) Bloch, C. and Ranganathan, R., 1993, Process-based cost modelling, *Microelectronics and Reliability*, Vol.33, No. 9, pp.1428.

10) Bodner, D., McGinnis, L, 2002, A structured approach to simulation modeling of manufacturing systems, *Proceedings of the 2002 Industrial Engineering conference.*

11) Bonney, M.C., Schofield, N.A., 1971, Computerised work study using the SAMMIE/AUTOMAT system, *International Journal Of Production Research*, Vol.9, No.3, pp.321-336.

12) Chan, D.S.K., 2003, Simulation modelling in virtual manufacturing analysis for integrated product and process design, *Assembly Automation*, Vol.23, No.1, pp.69-74.

13) Choi, C.K., and Ip, W.H., 1999, A comparison of MTM and RTM, *Work Study*, Vol. 48, No.2, pp.57-61.

14) Christmasson, M., Falck, A.C. et al., 2000, Modified method time measurements for ergonomic planning of production systems in the manufacturing industry, *International Journal Of Production Research*, Vol. 38, No.17, pp.4051-4059.

15) Creese, R.C., and Patrawala, T.B., 1999, A cost model for pultrusion process, *AACE International Transactions*, 1999, Morgantown, pp.EST.11.1-8.

16) Curran, R., Crosby, S., Butterfield, J. et al, 2005, Digital integrated manufacturing in collaborative engineering for reduced life cycle cost. In: J.X. Gao, Baxter, D., Sackett, P. (Editor), ICMR 2005:*3rd International conference on manufacturing research incorporating the 21st national NCMR*, Cranfield Univeristy.UK.

17) Dangelmaier, W., Fischer, M., et al, 2005, Virtual and augmented reality support for discrete manufacturing system simulation, *Computers in Industry*, Vol.56, pp.371-383.

18) Davis, J.J., 1998, Parametric estimates made easy, *AACE international Transaction*, ABI/INFORM Global, Morgantown, pp. Est.01.1-5.

19) Fagade, A., and Kazmer, D.O., 1999. Effect of Complexity on Cost & Time to Market of Injection Molded Parts, Proceeding of the Society of Plastics Engineers Annual Technical Conference, New York, NY.

20) Fagade, A., and Kazmer, D.O, 2000, Early Cost Estimation for Injection Molded Parts, *Journal of Injection Molding Technology*, Vol.4, No.3, pp.97-106.

21) Forcht, K., Cochran, k., 1999, Using data mining and data warehousing techniques, *Industrial management & Data Systems*, Vol.99, No.5, pp.198-196.

22) Gopalakrishnan, B., and Pathak, M.A., 1992, Machine parameter selection and cost estimation techniques: Applications in concurrent engineering, *ASME-Manufacturing International*, pp.249-256.

23) Hamilton, A.C., Richard, E. et al., 2002, Cost estimating best practices, *2002 AACE International Transactions*, pp. Est.02.1-5.

24) Han, Y., Zhou. C., McGinnis, L., et al, 2003, Paint Line color change reduction in automobile assembly through simulation, In: P.J.S.Chick, D.Ferrin, and D.J. Morrice, eds (Editor), *Proceedings of the 2003 Winter Simulation Conference*, Georgia Institute of Technology, Atlanta, U.S.A., pp. 1204-1209.

25) Harrell, C. and Tumay, K., 1997, Simulation made easy, *IIE Solutions*, Vol.29, No.7, pp.39(3).

26) Hart, D.F., Koster, G.A., 1997, Paint specification at GM, *Automotive Engineering*, Vol.105, No.6, pp.35-36.

27) Holes, L.G., Thomas, R., 1982, General Purpose Cost Modelling, E & FN Spon, London.

28) Hollmann, J.K., 1994, A parametric building cost estimating system, *AACE international Transactions*, Morgantown, pp. Est.4.1-7.

29) Hollocks, B.W., 1995, The impact of simulation in manufacturing decision making, *Control Engineering Practice*, Vol.3, No.1, pp.05-112.

30) Hoover, S., Perry, R., 1989, Simulation- A Problem-Solving Approach, Addison-Wesley, Reading, MA.

31) Jain, S., Choong, F.N., Aye, M.K. and Luo, M., 2001, Virtual Factory: an integrated approach to manufacturing systems modelling, *International Journal of Operations & Production Management*, Vol.21, No.5/6, pp.594-608.

32) Jiao, J. and Mitchell, M.T, 1999, A pragmatic approach to product costing based on standard time estimation, *International Journal of Operation & Production Management*, Vol.19, No.7, pp.738-755.

33) Jones, A. and Iuliano, M., 1997, A virtual manufacturing test bed, 1997, *IEEE International Conferencoen Systems, Man, And Cybernetics Computational Cybernetics And Simulation*, Florida, USA, October 12-15.

34) Joslin, W.L., 1976, 4M Computerized MTM-1, *Society of Manufacturing Engineer*, MS76-239, pp.15.

35) Kasravi, K., 1998, Paint Defect Control Using Data Mining, *Presented at MANUFACTURING 98, Published by Society of Manufacturing Engineers*, Dearborn, MI, Chicago, Illinois, pp. FC98-270-1-9, September 9-16, 1998.

36) Kelton, W.D., 2000, Experimental design for simulation, *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds., pp. 32-38.

37) Kirchain, R., and Field, F., 1997, Manufacturing cost estimation of large processing systems in: H.Y. Shon, Evans, J.W. and Apelian, D. (Editor), Proceedings of the Julian Szekely Memorial Symposium on Materails Processing. The Minerals, Metals & Materials Society, pp. 653-668.

38) Klingstam, P. and Gullander, P., 1999, Overview of Simulation tools for computer-aided production engineering, *Computers in Industry*, Vol.38, pp.173-186.

39) Kwong, C.K. and Bai, H., 2005, Fuzzy regression approach to process modeling and optimization of epoxy dispensing, *International Journal of Production Research*, Vol.43, No.12, pp.2359.

40) Law, M.E., 2002, Process modelling for future technologies, *IBM Journal of Research and Development*, Vol.46, No.2/3, pp.339-346.

41) Layer, A., Brinke E.T., Van Houten F., Kals H., and Haasis S., 2002, Recent and future trends in cost estimation, *International Journal of Computer Integrated Manufacturing*, Vol.15, No.6, pp.499-510.

42) Lederer, A.L. and Prasad, J., 1993, Systems development and cost estimating challenges and guidelines, *Information Systems Management*, 1993(Fall), pp.37-41.

43) Lehtonen, J.M. and Seppala, U., 1997, A methodology for data gathering and analysis in a logistic simulation project, *Integrated Manufacturing Systems*, Vol.8, No.6, pp.351-358.

44) Lenz, J. and Neitzel, R., 1995, Cost modelling: an effective means to compare alternatives, *Industrial Engineering*, Vol.27, No.1, pp.18-19.

45) McLean, C., 1998, Modelling production systems using virtual reality techniques, *Proceedings of IEEE Systems, Man and Cybernetics Conference*, San Diego, CA, pp. 344-347.

46) McNeill, M.F. and Thro, E., 1994, Fuzzy Logic; A practical approach, AP Professional, London, pp.13-18.

47) Moeller, A., 2004, Continuous simulation in material flow networks, *International Environmental Modelling and Software Society IEMSs 2004 International Conference*, 14-17 June 2004 University of Osnabrück, Germany.

48) Moore, L.T. and Creese, R.C., 1990, Manufacturing cost estimation, *Cost Engineering*, Vol.32, No.5, pp.17-21.

49) Mujtabi M.S., 1994, Simulation of modelling of manufacturing enterprise with complex material, information and control flows, *International Journal of Computer Integrated Manufacturing*, Vol.7, No.1, pp.29-46.

50) Needy, K.L., Nachtmann, H., Roztocki, N., et al., 2003, Implementing activity-based costing systems in small manufacturing firms: a field study, *Engineering Management Journal*, Vol.15, No.1, pp.3-10.

51) Orady, E.A. and Osman, T.A, 1997a, Capability study of robotics and manufacturing cell simulation software, *Computers & Industrial Engineering*, Vol.33, No.1-2, pp.83-86.

52) Orady, E.A. and Osman, T.A, 1997b, Virtual Reality software for robotics and manufacturing cell simulation, *Computers & Industrial Engineering*, Vol.33, No.1-2, pp.87-90.

53) Owen, J.V., 1997, Virtual Manufacturing, Manufacturing Engineering, Vol.119, No.4, pp.84-90.

54) Page, T., 2005, A classification of data mining techniques, *Society of Manufacturing Engineers*, TP05PUB16-1-10.

55) Paul, R.J. and Hlupic, V., 1994, Modelling of an automated manufacturing system using activity cycle diagrams, *International Journal Manufacturing System Design Technology*, Vol.1, No.2, pp.119-128.

56) Perara, T. and Liyanage, K., 1998, Rapid data collection in the simulation of manufacturing systems, *Proceedings of the Fourteenth National Conference on Manufacturing Research*, professional engineering publication Ltd, UK., University of Derby, UK, pp. 381-389.

57) Perez, A., Garcia-Fornieles, J.M., et al, 2000, The development of manufacturing cost-estimation processes for aerospace modification in: R. Perryman, and Ellis, C.D. (Editor), *Proceedings of the Sixteenth National Conference on Manufacturing Research*, Advances in manufacturing technology- XIV, University of East London, U.K., pp. 415-419.

58) Persson, J.F., 2002, The impact of different levels of detail in manufacturing systems simulation models, *Robotics and Computer-Integrated Manufacturing*, Vol.18, pp.319-325.

59) Roberts, C.A. and Hermosillo, E.P., 2000, An automated machining cost estimator, *Engineering Valuation and Cost Analysis*, Vol.3, pp.27-42.

60) Robertson, N. and Perera, T., 2002, Automated data collection for simulation?, *Simulation Practice and Theory*, Vol.9, pp.349-364.

61) Ronald, J. and Hatwell, P.E., 1993, Cost engineering system for the future, *Cost Engineering*, Vol. 35, No.3, pp.13-17.

62) Roy, R., Rush, C.S., Tuer, G., 2002, Cost estimating rationale capture, *AACE International Transactions*, Morgantown, pp. Est.12.1-10.

63) Roy, R., Allen, D. and Zamore, O., 2004, Cost of photochemical machining, *Journal of Materials Processing Technology*, Vol.149, No.1-3, pp.460-465.

64) Sadowski, D.A. and Grabau, M.P., 1999, Tips for successful practice of simulation, *Proceedings of the 1999 Simulation Conference*, Phoenix, Arizona, December 5-8, 1999.

65) Samid, G.P., 1998, Top-Down (or Narrow-Down) Cost Estimation, *AACE International Transactions*, pp. Est.04.1-4.

66) Shukla, C., Vazquez, M. and Chen, F.F., 1996, Virtual Manufacturing: An Overview, Computers Industrial Engineering, Vol.31, No.1/2, pp.79-82.

67) Skitmore, M. and Marston, V., 1999, Cost Modelling, E & FN Spon, London.

68) Sonmez, R., 2005, Review of conceptual cost modeling techniques, *AACE International Transactions*, ES71.

69) Thomson, N., 1995, Simulation in Manufacturing. John Wiley & Sons Inc., Chichester.

70) Ulgen, O.M., Shore, J., Coffman, G., et al, 1999, Increasing the power and value of manufacturing simulation via collaboration with other analytical tools: a panel discussion, *Proceedings of the 1999 Winter Simulation Conference*, Dearborn, MI, U.S.A., pp.749-753.

71) Vincze, M., Pichler, A., Biegelbauer, G. and Hausler, K., 2002, Automatic robotic spray painting of low volume high variant parts, *Proceedings of the 33rd ISR (International Symposium on Robotics)* October 7-11, 2002, pp. 6.

72) Wreden, N., 1998, The future of process Modeling, *Information Week*, pp. 6A.

73) Wygant, R.M. and Donaghey, C.E., 1986, A computer model to compare robot and human performance, Robotics systems education and training. SME; MS86-521, Detroit, Michigan, pp. 1-11.

74) Zakarian, A. and Kusiak, A, 2000, Analysis of Process Models, *IEEE Transactions on Electronics Packaging Manufacturing*, Vol.23, No.2, pp.137-147.

75) Zhao, J., Cheung, W.M., and Young, R.I.M., 1999, A consistent manufacturing data model to support virtual enterprises, *International Journal of Agile Management*, Vol.1, No.3, pp.150-158.

Publications

Shaik, T, Stockton, D., and Khan, Z., 2005, Process Time Estimating Model Development using Virtual Process Models and Data Mining, *Proceedings of the 3rd Intl. Conference on Manufacturing Research (ICMR05), 6-8 Sept, Canfield University, U.K.*

Shaik, T, Steeple, D. and Stockton, D., 2004, A Review of the Cost Model Development Process, *Proceedings of the 2nd Intl Conference on Manufacturing Research (ICMR04), Sheffield Hallam University, U.K.*

Baguley, P., Shaik, T., Fresco, J., Stockton, D., 2003, Cost Modelling with Fuzzy Logic, *Proceedings of the20th Intl. Manufacturing Conference (IMC 20), Sept, University of Cork, Ireland.*

APPENDICES

Appendix 3.1A

Table 3.1A: Modified Process Scoping Characteristics of COSTMOD Methodology

	CS	DES	VM	MPM	PMTS	VDG
Time Required to develop initial data generation model (Weeks/Months)	BS	DDE	DDE	BS	BS	
Time required to carry out data generation experimentation (Weeks/Months)	BS	DDE	DDE	BS	BS	DNE
Time required to collect the generated data (Hours/Days/Weeks)	BS	DDE	DDE	BS	BS	DNE
Necessity to identify process or method experts for model development (Yes/No)	BS	DDE	DDE	BS	BS	

DDE- Discussion with Delmia's Expertise
DN-Discussion with Nouldus Expertise
BS- Brainstorming session with process model experts at DMU

Table 3.1B: Comparisons of Data Generation Methods (I)

		VM	DES	CS	MPM	PTMS	VDG
Development state of process	Concept	(Alabastro, 1995)		(Baines, 1999)		(Barrett, 1997)	(Nouldus 2004)
	Detail Design	(Klingstam, 1999; Orady, 1997)	(Kibira, 2002)	(Baines, 1999)	(Wood 1999)	(Clark, 1977; Cohen, 1998)	(Nouldus 2004)
	Prototype	(Anonymous, 1993; CIMdata, 2003) (Orady, 1997)	(Kibira, 2002)	(Baines, 1999)	(Wood 1999)	(Clark, 1977)	(Nouldus 2004)
	Commercial		(Williams, 107)	(Baines, 1999)	Acosta	(Barrett, 1997)	(Nouldus 2004)
	New Product/Process	(CIMdata, 2003) 119	(Williams, 107)	(Baines, 1999)	Perkins(1996)	(Al-Dohaim, 1993)	(Nouldus 2004)
	Modified Product/Process	(CIMdata, 2003; Orady, 1997)	(Williams, 107)	(Baines, 1999)	Perkins(1996)	(Clark, 1977)	(Nouldus 2004)
Production Volume	Low Volume/One Off	(Worn 2000)	(Ulgen, 2000)	X			(Nouldus 2004)
	Medium Volume	(Worn 2000)		X			(Nouldus 2004)
	High Volume	(Worn 2000)		X			(Nouldus 2004)
Production Type	Project	(Lee, 2001)	(Heilala, 1999)	X			(Nouldus 2004)
	Job Shop	(CIMdata, 2003)	(Heilala, 1999)	X			(Nouldus 2004)
	Batch	(CIMdata, 2003)	(Klingstam, 1999)	X	(Petropolakis 1998)		(Nouldus 2004)
	Flow	(CIMdata, 2003)	(Klingstam, 1999)	X	(Petropolakis 1998)		(Nouldus 2004)
	Continuous	(Klingstam, 1999; Orady, 1997)		Gorlani (1993)	(Petropolakis 1998)		(Nouldus 2004)

Table 3.1C: Comparisons of Data Generation Methods (II)

		VE	DES	CS	MPM	PTMS	VDG
Resource able to be estimated	Direct Material Cost						
	Indirect Material Cost						
	Direct Equipment Cost						
	Indirect Equipment Cost						
	Direct Labour Cost						
	Indirect Labour Cost						
	Direct Process Time	(Brown, 2004)	(Hollocks, 1995)	Switek(1997)	(Matko, 1992)	(Wygant, 1986)	(Nouldus 2004)
	Indirect Process Time	(Brown, 2004)	(Hollocks, 1995)	Switek(1997)		(Wygant, 1986)	(Nouldus 2004)
Product Feature levels	Product Level	(Sung, 2003)	(Kendall, 1998)				(Nouldus 2004)
	Component Level	(Brown, 2004)	(Kendall, 1998)	(Kilgore, 1997)			(Nouldus 2004)
	Component Feature Level	(Waurzyniak, 2004)	(DeVin 2004)				(Nouldus 2004)
Process Feature levels	Machine Level	(Sung, 2003) 75	Rohrer (weblink)		(Matko, 1992)	(Anna 1995)	(Nouldus 2004)
	Machine Assembly Level	(Waurzyniak, 2004) (Brown, 2004)		(Baines, 1999)			(Nouldus 2004)
	Machine Sub-Assembly Level	(Waurzyniak, 2004) 169					(Nouldus 2004)
Process Activity levels	Process Level	(Sung, 2003)			(Matko, 1992)		(Nouldus 2004)
	Process Operation Level	(Waurzyniak, 2004)	Rohrer		(Matko, 1992)	(Anna 1995)	(Nouldus 2004)
	Operational Activity Level	(Brown, 2004; Waurzyniak, 2004)	(DeVin 2004)	(Baines, 1999)		(Kapoor 1990)	(Nouldus 2004)

Appendix 4.1A

Table 4. 1A: Milling Process OA Variables Levels for Model-1

Process Variables	Levels		
	Level 1	Level 2	Level 3
Tool Diameter (D)	3.2	6.7	12.8
Depth of Cut (Dc)	1.9	3.8	7.6
Machining Length (Lc)	115	396	680
Surface speed (Vc)	19	25	31
Number of Teeth (z)	3	3	3

Table 4.1B: L27 (5^3) Orthogonal Array for Milling Experiments

Exp.No	Lc	D	Dc	z	Vc
1	115	3.2	1.9	3	19
2	115	3.2	1.9	3	25
3	115	3.2	1.9	3	31
4	115	6.7	3.8	3	19
5	115	6.7	3.8	3	25
6	115	6.7	3.8	3	31
7	115	12.8	7.6	3	19
8	115	12.8	7.6	3	25
9	115	12.8	7.6	3	31
10	396	3.2	3.8	3	19
11	396	3.2	3.8	3	25
12	396	3.2	3.8	3	31
13	396	6.7	7.6	3	19
14	396	6.7	7.6	3	25
15	396	6.7	7.6	3	31
16	396	12.8	1.9	3	19
17	396	12.8	1.9	3	25
18	396	12.8	1.9	3	31
19	680	3.2	7.6	3	19
20	680	3.2	7.6	3	25
21	680	3.2	7.6	3	31
22	680	6.7	1.9	3	19
23	680	6.7	1.9	3	25
24	680	6.7	1.9	3	31
25	680	12.8	3.8	3	19
26	680	12.8	3.8	3	25
27	680	12.8	3.8	3	31

Table 4. 1C: Milling Process OA Variables Levels for Model-3

Process Variables	Levels		
	Level 1	Level 2	Level 3
Tool Diameter (D)	25.4	38.1	50.8
Depth of Cut (Dc)	15.2	25.4	38.1
Machining Length (Lc)	381	762	1524
Surface speed (Vc)	152	228	304
Number of Teeth (z)	4	4	5

Table 4.1D: L27 (5^3) Orthogonal Array for Milling Experiments

Exp. No.	Lc	D	Dc	z	Vc
1	762	12.8	7.7	3	76
2	762	12.8	7.7	3	91
3	762	12.8	7.7	3	106
4	762	25.4	12.8	3	76
5	762	25.4	12.8	3	91
6	762	25.4	12.8	3	106
7	762	38.1	19.05	4	76
8	762	38.1	19.05	4	91
9	762	38.1	19.05	4	106
10	1524	12.8	12.8	3	76
11	1524	12.8	12.8	3	91
12	1524	12.8	12.8	3	106
13	1524	25.4	19.05	3	76
14	1524	25.4	19.05	3	91
15	1524	25.4	19.05	3	106
16	1524	38.1	7.7	4	76
17	1524	38.1	7.7	4	91
18	1524	38.1	7.7	4	106
19	2032	12.8	19.05	3	76
20	2032	12.8	19.05	3	91
21	2032	12.8	19.05	3	106
22	2032	25.4	7.7	3	76
23	2032	25.4	7.7	3	91
24	2032	25.4	7.7	3	106
25	2032	38.1	12.8	4	76
26	2032	38.1	12.8	4	91
27	2032	38.1	12.8	4	106

Appendix 4.2A

Table 4. 2A: Spray Paint Variable Levels for Model-2

Process Variable	Levels		
	Level 1	Level 2	Level 3
Distance from Work piece(D_{wp}-mm)	200	250	325
Paint flow rate (P_{fr-} cc/min)	350	450	550
Paint Gun Speed (G_s-mm/sec)	300	400	500
Paint Gun range (G_{rg} -mm)	300	350	400
Path Length (P_{tl}-mm)	6202	6982	7505

Table 4.2B: L27 (5^3) Orthogonal Array for Model 2 Spray Paint Experiments

Exp.No	D_{wp}	P_{fr}	G_s	G_{rg}	P_{tl}
1	200	350	300	300	6202
2	200	350	300	300	6982
3	200	350	300	300	7505
4	200	450	400	350	6202
5	200	450	400	350	6982
6	200	450	400	350	7505
7	200	550	500	400	6202
8	200	550	500	400	6982
9	200	550	500	400	7505
10	250	350	400	400	6202
11	250	350	400	400	6982
12	250	350	400	400	7505
13	250	450	500	300	6202
14	250	450	500	300	6982
15	250	450	500	300	7505
16	250	550	300	350	6202
17	250	550	300	350	6982
18	250	550	300	350	7505
19	325	350	500	350	6202
20	325	350	500	350	6982
21	325	350	500	350	7505
22	325	450	300	400	6202
23	325	450	300	400	6982
24	325	450	300	400	7505
25	325	550	400	300	6202
26	325	550	400	300	6982
27	325	550	400	300	7505

Table 4.2C: Spray Paint Variable Levels for Model-3

Process Variable	Levels		
	Level 1	Level 2	Level 3
Distance from Work piece(D_{wp}-mm)	225	250	300
Paint flow rate (P_{fr}- cc/min)	400	500	600
Paint Gun Speed (G_s-mm/sec)	450	550	650
Paint Gun range (G_{rg} -mm)	325	350	375
Path Length (P_{tl}-mm)	4216	4950	5521

Table 4.2D: L27 (5^3) Orthogonal Array for Model 3 Spray Paint Experiments

Exp.No	D_{wp}	P_{fr}	G_s	G_{rg}	P_{tl}
1	225	400	450	325	4216
2	225	400	450	325	4950
3	225	400	450	325	5521
4	225	500	550	350	4216
5	225	500	550	350	4950
6	225	500	550	350	5521
7	225	600	650	375	4216
8	225	600	650	375	4950
9	225	600	650	375	5521
10	250	400	550	375	4216
11	250	400	550	375	4950
12	250	400	550	375	5521
13	250	500	650	325	4216
14	250	500	650	325	4950
15	250	500	650	325	5521
16	250	600	450	350	4216
17	250	600	450	350	4950
18	250	600	450	350	5521
19	300	400	650	350	4216
20	300	400	650	350	4950
21	300	400	650	350	5521
22	300	500	450	375	4216
23	300	500	450	375	4950
24	300	500	450	375	5521
25	300	600	550	325	4216
26	300	600	550	325	4950
27	300	600	550	325	5521

Table 4.2E: Spray Paint Variable Levels for Model-4

Process Variable	Levels		
	Level 1	Level 2	Level 3
Distance from Work piece(D_{wp}-mm)	200	250	325
Paint flow rate (P_{fr}- cc/min)	200	400	600
Paint Gun Speed (G_s-mm/sec)	300	500	700
Paint Gun range (G_{rg} -mm)	350	450	550
Path Length (P_{tl}-mm)	5434	6490	7892

Table 4.2F: L27 (5^3) Orthogonal Array for Model 4 Spray Paint Experiments

Exp. No	D_{wp}	P_{fr}	G_s	G_{rg}	P_{tl}
1	200	200	300	350	5434
2	200	200	300	350	6490
3	200	200	300	350	7892
4	200	400	500	450	5434
5	200	400	500	450	6490
6	200	400	500	450	7892
7	200	600	700	550	5434
8	200	600	700	550	6490
9	200	600	700	550	7892
10	250	200	500	550	5434
11	250	200	500	550	6490
12	250	200	500	550	7892
13	250	400	700	350	5434
14	250	400	700	350	6490
15	250	400	700	350	7892
16	250	600	300	450	5434
17	250	600	300	450	6490
18	250	600	300	450	7892
19	325	200	700	450	5434
20	325	200	700	450	6490
21	325	200	700	450	7892
22	325	400	300	550	5434
23	325	400	300	550	6490
24	325	400	300	550	7892
25	325	600	500	350	5434
26	325	600	500	350	6490
27	325	600	500	350	7892

Appendix 5.1A

Experimental Results for DS_2

From Tables 5.1, 5.10 and 5.14, it is evident that FD has split the data set in to two i.e. DS_1 and DS_2. DS_1 is the data set with data points, which are explored by FD algorithms and are found to be most influencing on the resulting dependent variable. Where as DS2 are the remaining data points from the full data set, i.e. those data point which are not in the distribution of FD. As mentioned in section 6.4.1, that these data points may or may not be helpful in the analysis. Therefore, in order to identify the impact of these data points on cycle time predictive data mining is carried out for each of these data set (DS_2) of individual processes under consideration.

5.1A Vertical End Milling

In this Appendix DS_2 for Milling was explored using three data mining algorithms i.e. SLR, FL and PnP.

1) Stepwise Linear Regression (SLR)

The following model was developed using the Stepwise Linear Regression Algorithm, i.e.:

$$T_m = 22.5584 - 10.9127 \cdot D_c - 0.0188828 \cdot V_c - 0.00143107 \cdot n + 0.319402 \cdot L_c \qquad \textbf{(1a)}$$

Where:

T_m	=	Machining time (min)
D_c	=	Depth of cut (mm)
V_c	=	Surface speed of tool (m/min)
n	=	Spindle speed (rpm)
Lc	=	Length of machining (mm)

Removal of those terms that contribute small amounts to the values of Tm produces a simplified version of Equation (1), i.e. Equation (2):

$$T_m = 22.55 - 10.91 \cdot D_c - 0.02 \cdot V_c - 0.001 \cdot n + 0.32 \cdot L_c \qquad (2a)$$

The statistical values shown in Table 5.1A indicate the estimating accuracy of Equation (2a).

Table 5. 1A: Estimating Accuracy of Equation (2a)

R-Squared (R^2)	0.11
Standard Error	0.96
Standard Deviation	40.42

2) Find Laws Algorithm (FL)

The following model for estimating T_m was developed using the Find Laws Algorithm, i.e.

$$T_m = (0.62078 \cdot F_t \cdot L_c)(F_t \cdot V_f \cdot D_c - 0.00135935 \cdot D_c + 3.62728e\text{-}005)^{-1} \qquad \textbf{(3a)}$$

By removal of insignificant terms the simplified version of Equation 3 is as follows:

$$T_m = (0.62 \cdot F_t \cdot L_c)(D_c(F_t \cdot V_f - 0.001))^{-1} \qquad \textbf{(4a)}$$

Where:
- T_m = Machining time (min)
- L_c = Machining length (mm)
- V_f = Feed rate (mm/min)
- F_t = Feed per tooth (mm)
- Dc = Depth of cut (mm)

Table 5.1B shows the estimating accuracy of Equation (4a) obtained using the Find Laws algorithm.

Table 5. 1B: Estimating Accuracy of Equation (4a)

R-Squared (R^2)	0.94
Standard Error	0.23
Standard Deviation	9.77

3) PolyNet Predictor Algorithms (PnP)

As PolyNet predictor is an Artificial Neural Network based algorithm when trained which represents a hierarchical network that comprises of nodes and layers in the form of the hidden relationships found between the dependent and independent variables. However, it does not display the results in a compact and readable form for large neural networks, which is one of the limitations of artificial neural network (Polyanalyst). The network outputed from the PolyNet predictor for the milling DS_2 process consists of 2 layers and 7 nodes.

Table 5.1C lists the Mean Absolute Percentage Error (MAPE) obtained for the milling cost models through use of Equations (2a), (4a) and (5a).

Table 5. 1C: MAPE for Milling Models

Algorithms	MAPE
Stepwise Linear Regression	1409.65
Find Laws	296.85
PolyNet Predictor	2826.62

Appendix 5.2A

Automated Spray Paint

In this Appendix DS_2 for Automated Spray Paint was developed using data mining algorithms i.e. SLR, FL and PnP.

1) Stepwise Linear Regression

Equation (5a) for Automated Paint Spraying for DS_2 was developed using Stepwise Linear Regression;

$$C_{tm} = 5.41342 - 0.004082 \cdot P_{fr} - 0.0048497 \cdot G_s + 0.0012524 \cdot P_{tl} + 0.054981 \cdot P_{ns} \qquad \textbf{(5a)}$$

A simplified version of Equation (5a) was developed by removing non-significant terms, i.e. Equation (6a):

$$C_{tm} = 5.41 - 0.004 \cdot P_{fr} - 0.005 \cdot G_s + 0.001 \cdot P_{tl} + 0.05 \cdot P_{ns} \qquad \textbf{(6a)}$$

Where:

C_{tm} = Painting cycle time,

P_{fr} = Paint fluid flow rate,

G_s = Paint gun speed,

P_{tl} = Path length, and

P_{ns} = Paint sprayed.

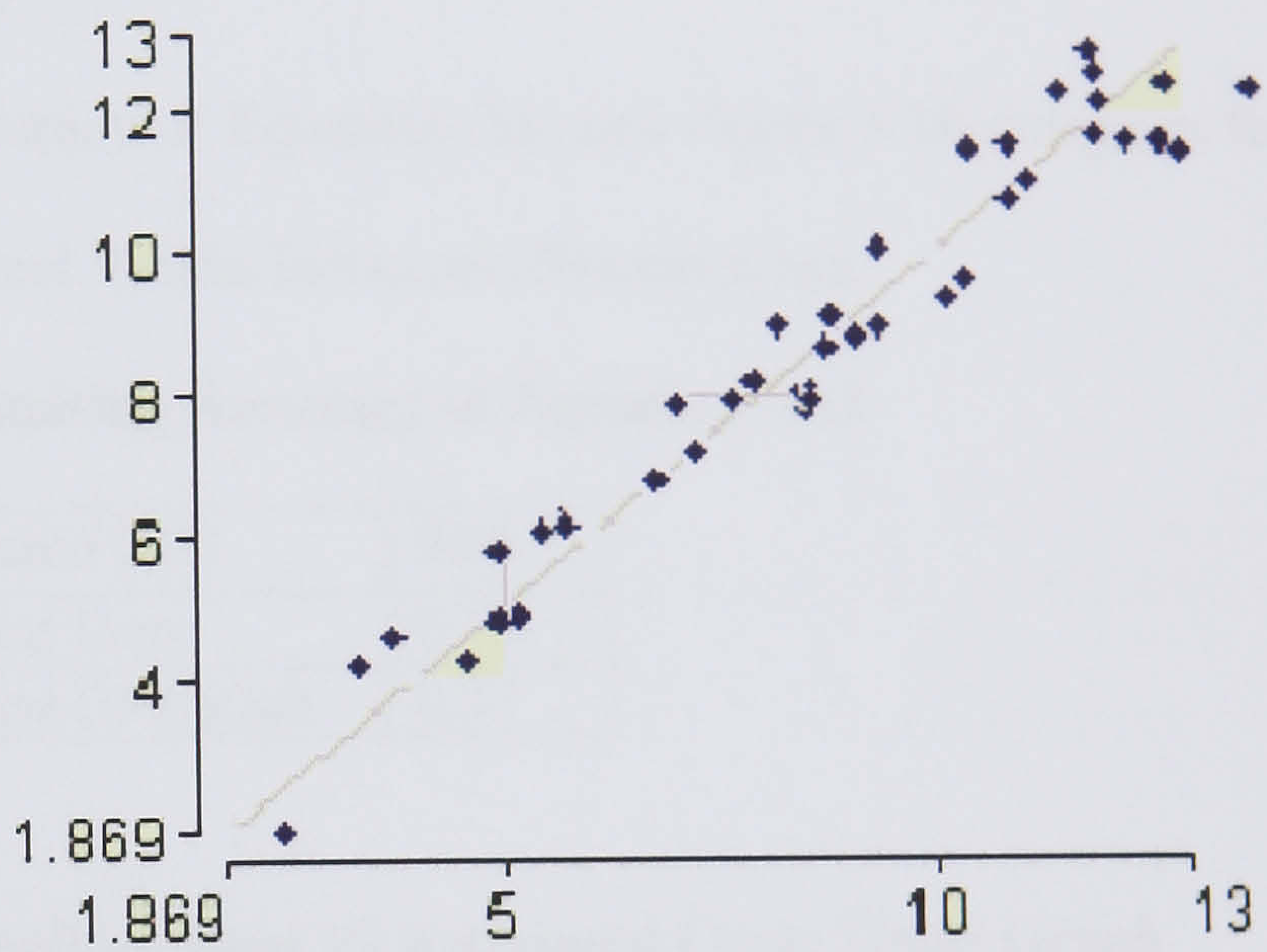
Table 5.2A reflects the measure of accuracy and significance of the model developed and Figure 5.2A compares for Automated Paint spraying Real Versus Estimated Process times plotted using Equation (6a).

Table 5. 2A: Estimating Accuracy of Equation (6)

R-Squared (R^2)	0.95
Standard Error	0.21
Standard Deviation	0.63

Figure 5.2A: Stepwise Linear Regression- Real Vs Estimated Cycle Times

Estimated Process Times (C_{tm}) using Equation (6a)



Real Process Times (C_{tm})

2) Find Laws

The following model for Automated Paint Spraying was developed using the Find Laws Algorithm,

$$C_{tm} = (0.0016157 \cdot P_{fr} \cdot P_{tl} - 0.69669 \cdot P_{tl} + 0.04798 \cdot P_{ns} \cdot P_{fr} - 26.2554 \cdot P_{ns})(P_{fr} - 458.515)^{-1} \quad \textbf{(7a)}$$

The simplified version of the Equation (7a) was developed by removing non-significant terms, i.e. Equation (8a):

$$C_{tm} = (0.002 \cdot P_{fr} \cdot P_{tl} - 0.7 \cdot P_{tl} + 0.05 \cdot P_{ns} \cdot P_{fr} - 26.25 \cdot P_{ns})(P_{fr} - 458.5)^{-1} \quad \textbf{(8a)}$$

Where: C_{tm} = Process cycle time,

P_{fr} = Paint fluid flow rate,

P_{tl} = Path length, and

P_{ns} = Paint sprayed.

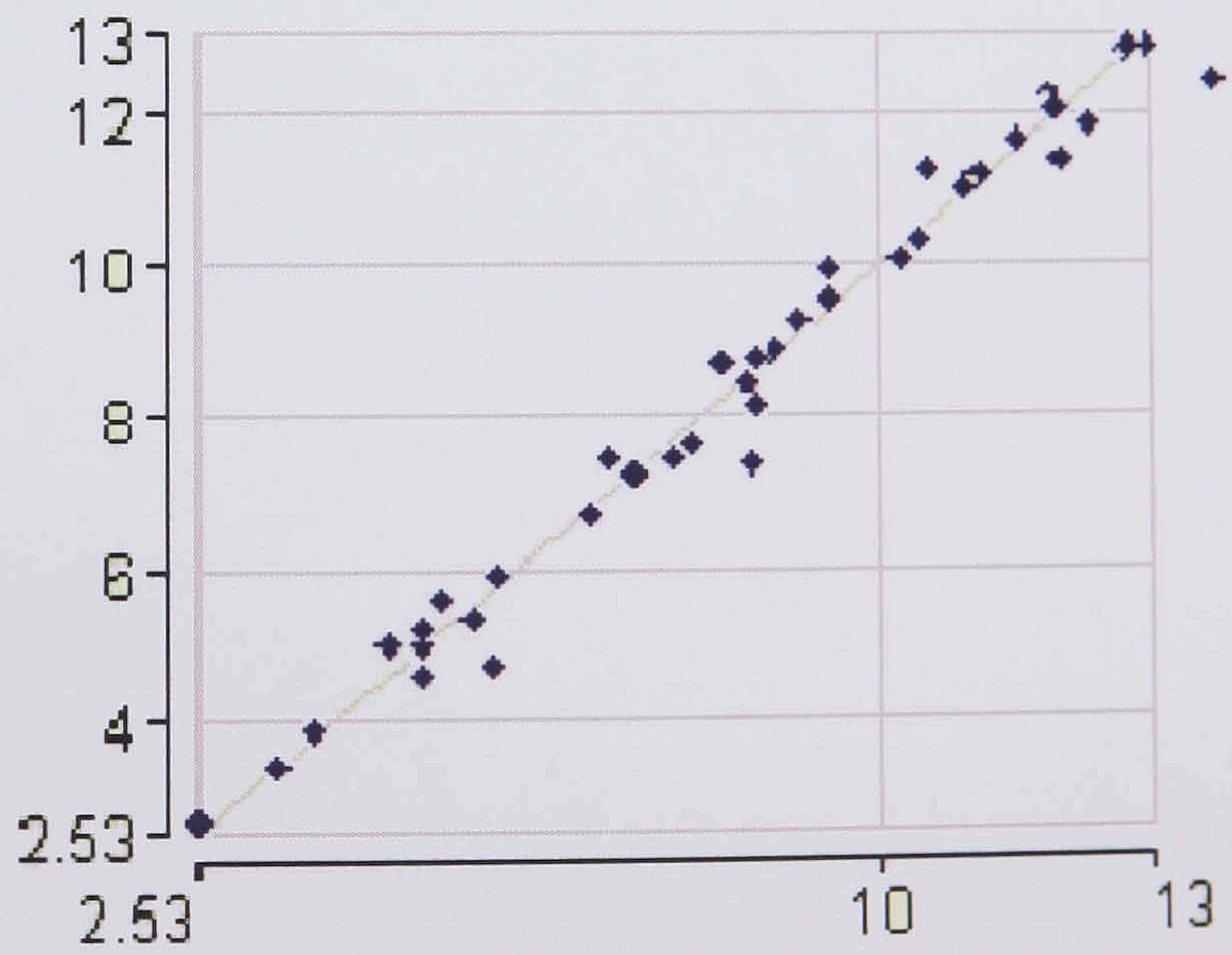
Table 5.2B shows the estimating accuracy of Equation (8a) and Figure 5.2B compares for Automated Spray Paint process the Real Versus Estimated Process times.

Table 5.2B: Estimating Accuracy of Equation (8a)

R-Squared (R^2)	0.98
Standard Error	0.12
Standard Deviation	0.37

Figure 5.2B: Find Laws Algorithm- Real Vs Estimated Cycle Times Graph

Estimated Process Times (C_{tm}) using Equation (8a)



Real Process Times

Table 5.2C lists the MAPE obtained for the automated spray painting cost models through use of Equations (6a) and (8a).

Table 5. 2C: MAPE for Automated Paint Spraying

Algorithms	MAPE
Stepwise Linear Regression	22.50
Find Laws	22.55
PolyNet Predictor	-

Appendix 5.3A

Turning

In this Appendix 5.3A, CERs for Turning DS_2 was developed using data mining algorithms i.e. SLR, FL and PnP.

1) Stepwise Linear Regression

Equation (9a) for Turning was developed using Stepwise Linear Regression:

$$\begin{aligned} T = {} & 9.41740 \cdot p_s - 26.4442 \cdot n + 0.0281831 \cdot t_{sa} + 14.9189 \cdot n_t \\ & + 0.13323 \cdot t_{sb} - 2.8292 \cdot B_s + 1.48171 \cdot t_{ln} + 6.96655 \cdot n_o + 6.2598 \cdot t_{pt} \end{aligned} \tag{9a}$$

A simplified version of Equation (9a) is provided by Equation (10a):

$$\begin{aligned} T = {} & 9.42 \cdot p_s - 26.44 \cdot n + 0.03 \cdot t_{sa} + 14.92 \cdot n_t \\ & + 0.13 \cdot t_{sb} - 2.82 \cdot B_s + 1.48 \cdot t_{ln} + 6.96 \cdot n_o + 6.26 \cdot t_{pt} \end{aligned} \tag{10a}$$

Where:

T = Cycle time,

r_e = Proportion of material removed by external machining,

t_{sa} = Basic set-up time for machine,

n_t = Number of tools,

t_{sb} = Set-up time per tool,

B_s = Batch size,

t_{ln} = Loading and unloading time,

n_o = Number of operations, and

t_{pt} = Tool positioning time per operation.

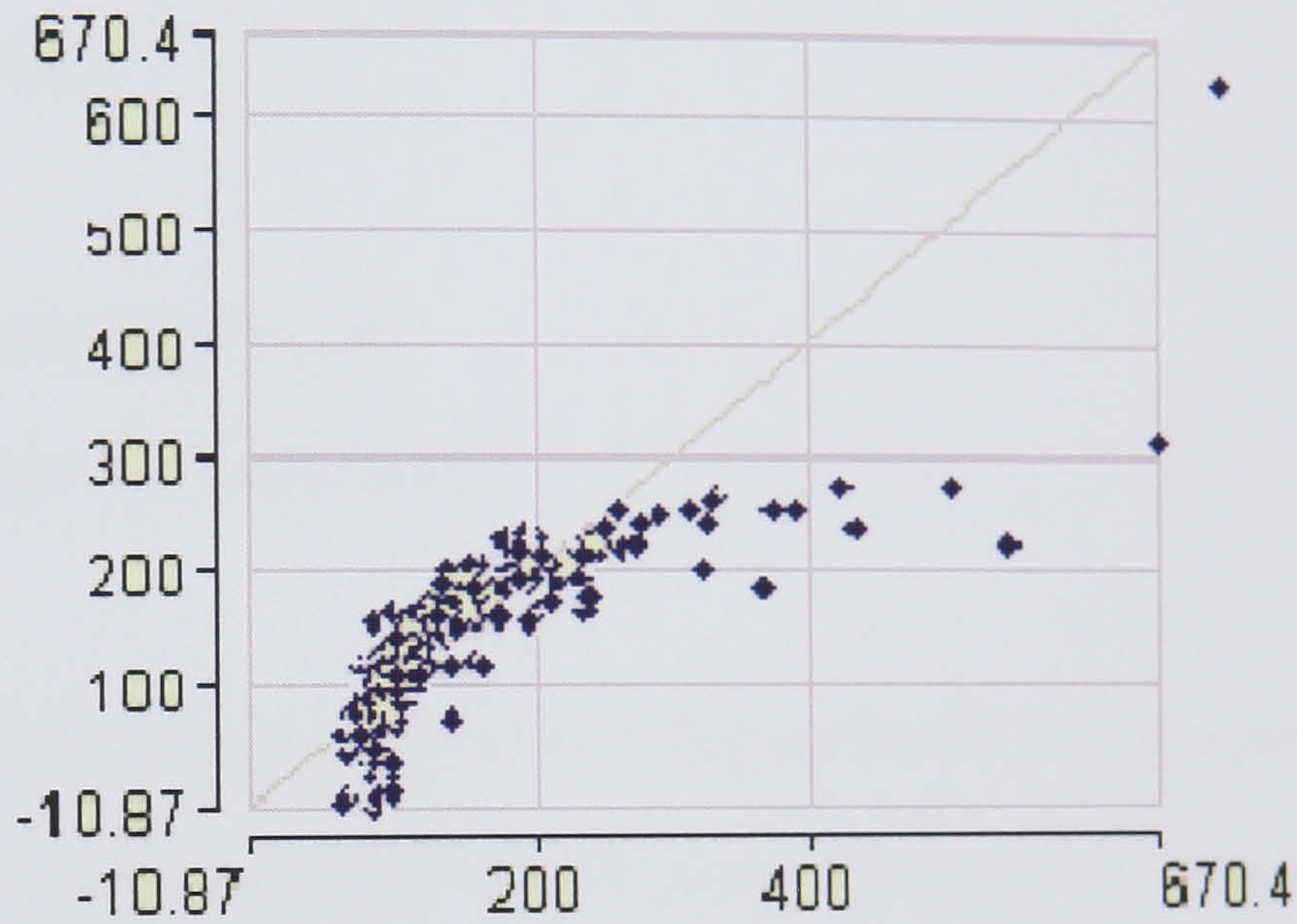
Table 5.3A provides the estimating accuracy of Equation (10a) and Figure 5.3A compares for the Turning operation the Real Versus Estimated Process time derived from Equation (10a).

Table 5. 3A: Estimating Accuracy of Equation (10a)

R-Squared (R^2)	0.57
Standard Error	0.65
Standard Deviation	49.31

Figure 5.3A: Stepwise Linear Regression –Real Vs Estimated Process Times

Estimated Process Times (*T*) using Equation (10a



Real Process Times

2) Find Laws

Equation (11a) for Turning was developed using the Find Laws Algorit

$$T = (74.3901 \cdot B_s + 538.544 - 4.1726 \cdot n_t^2 + 1.5849 \cdot t_{sa} + 0.05930 \cdot t_{sb} \cdot B_s + 0.22032 \cdot t_{sb} \cdot n_t^2)\,(1 + 1.51184 \cdot B_s - 0.00238728 \cdot t_{sb})^{-1} \quad \textbf{11a}$$

The simplified version of equation (11a) is shown in Equation (12a):

$$T=(538.54+74.39\cdot B_s-4.17\cdot n_t^2+1.58\cdot t_{sa}+0.06\cdot t_{sb}\cdot B_s+0.22\cdot t_{sb}\cdot n_t^2)$$
$$(1+1.51\cdot B_s-0.003\cdot t_{sb})^{-1} \qquad \textbf{(12a)}$$

Where:

T = Cycle time,

r_e = Proportion of material removed by external machining,

t_{sa} = Basic set-up time for machine,

n_t = Number of tools,

t_{sb} = Set-up time per tool,

B_s = Batch size,

t_{ln} = Loading and unloading time, and

t_{pt} = Tool positioning time per operation.

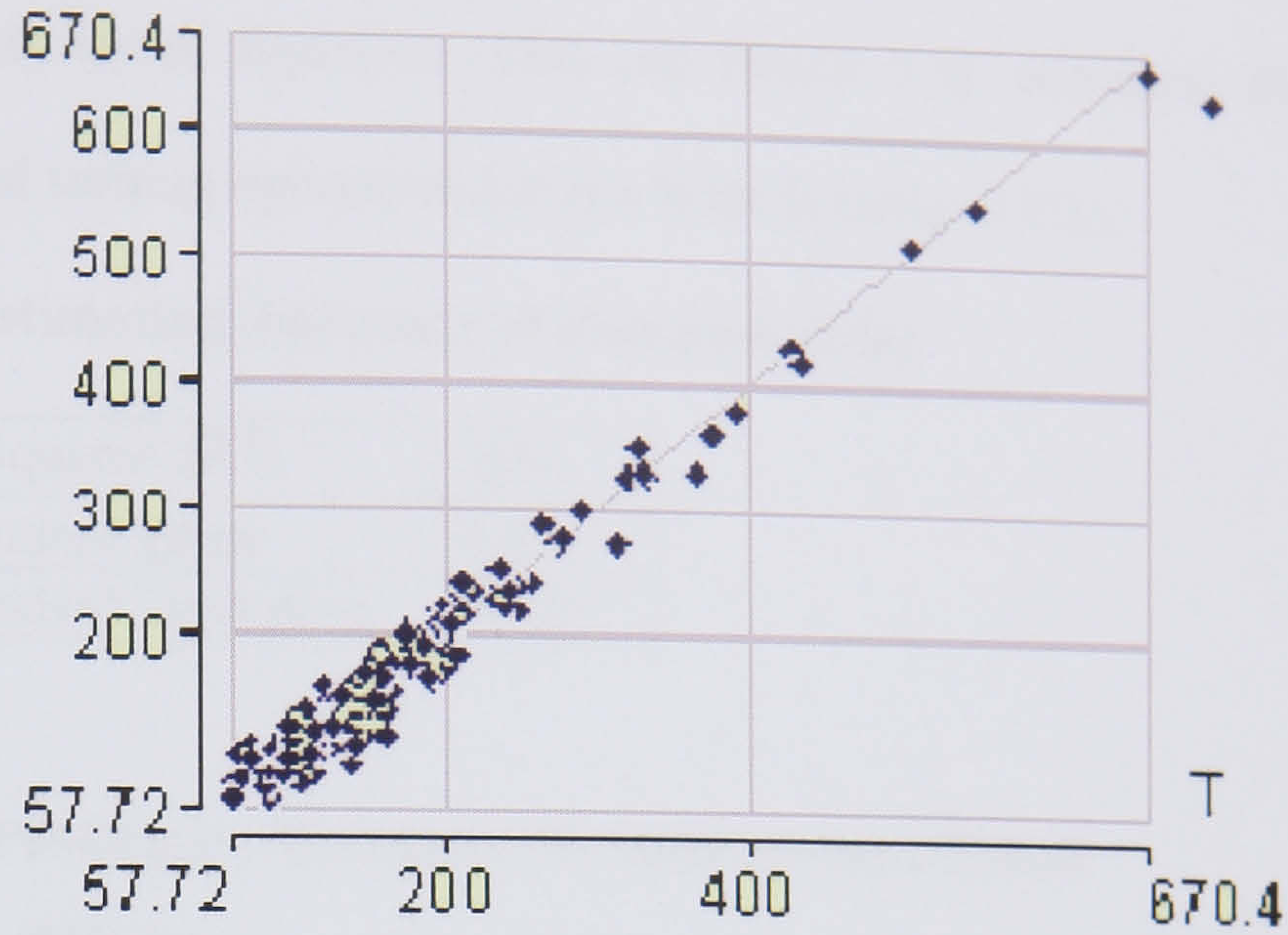
Table 5.3B indicates the estimating accuracy of Equation (12a) and Figure 5.3B compares for Turning operation the Real Versus Estimated Process times derived using Equation (12a).

Table 5.3B: Estimating Accuracy of Equation (12a)

R-Squared (R^2)	0.96
Standard Error	0.15
Standard Deviation	15.01

Figure 5.3B: Find Laws- Real Vs Estimated Process Times

Estimated Process Times (T) using Equation (12a)



Real Process Times

3) PolyNet Predictor

The following model for Turning was developed using the PolyNet Predictor Algorithm,

$$T = (376.45 + B_s(-15.80 + B_s \cdot (0.30 - 0.0017 \cdot B_s))) \quad \textbf{(13a)}$$

Where:

- T = Cycle time,
- n_t = Number of tools,
- t_{sb} = Set-up time per tool,
- B_s = Batch size, and
- r_i = Proportion of material removed by internal machining.

The architecture of Equation (13a) consists of 2 network layers and 7 network nodes. Table 5.3C indicates the estimating accuracy of Equation (13a) and Figure 5.3C compares the Estimated Vs Real Process times of turning operation derived from Equation (13a).

Table 5. 3C: Estimating Accuracy of Equation (13a)

R-Squared (R^2)	0.54
Standard Error	0.67
Standard Deviation	51.09

Figure 5.3C: PolyNet Predictor - Estimated Vs Real Process Times

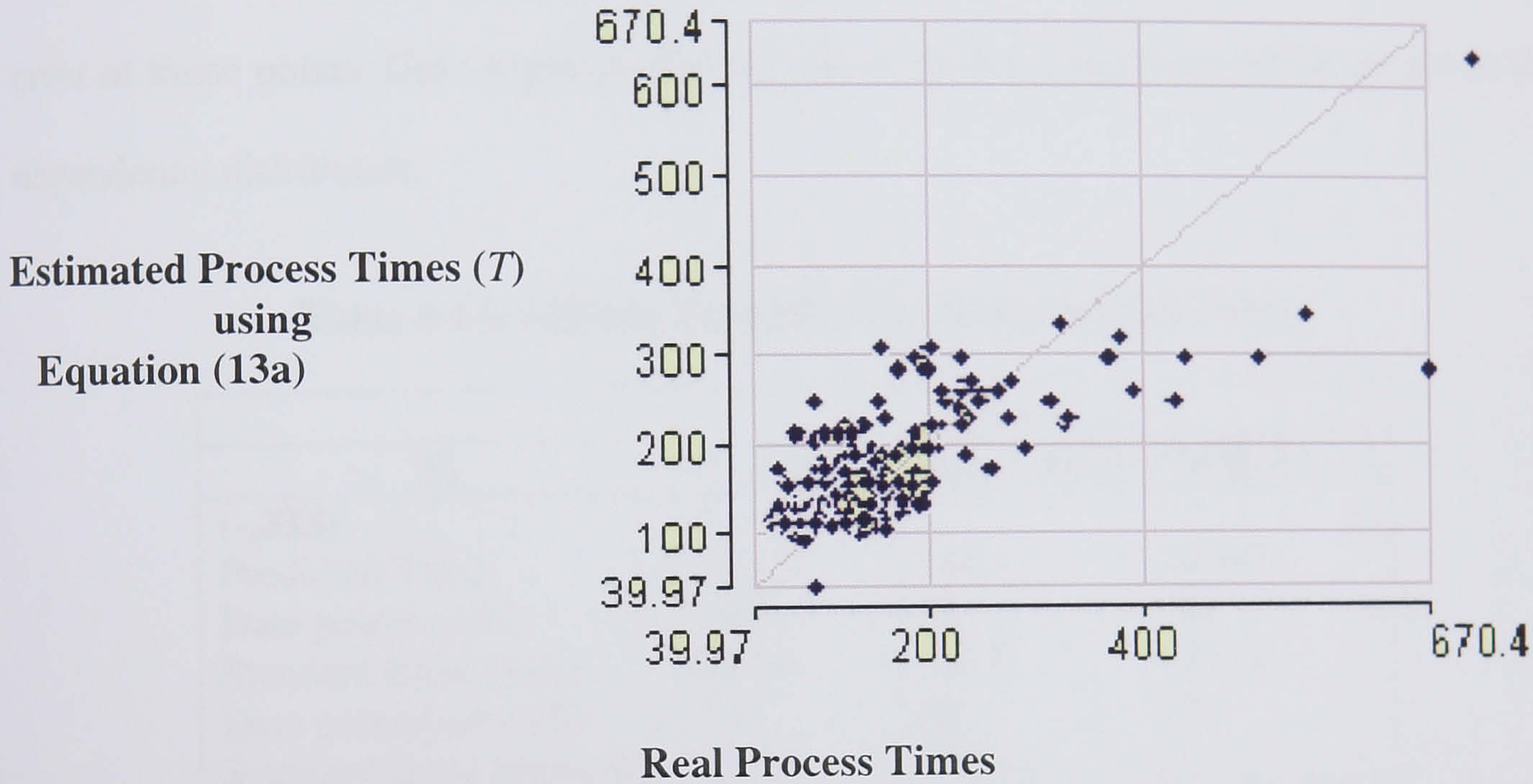


Table 5.3D lists the MAPE obtained for the turning cost models through use of Equations (10a), (12a) and (13a).

Table 5. 2: MAPE for Turning Process Models

Algorithms	MAPE
Stepwise Linear Regression	52.25
Find Laws	13.25
PolyNet Predictor	33.17

Appendix 6.1A

Vertical End Milling

Table 6.1A is the output of Find Dependencies is the dependencies table for Milling. The table divides the data into segments along two axes, these two axes represents the variables that had the strongest influence on the target dependent variables. Here horizontal axes represent Lc and Vertical axes represent Vf. Within each cell, five numbers are listed: the predicted value (Tm), the number of points in the cell, the standard error of those points, the number of points in the cell that fall into the dependent subpopulation, and the standard error of those points. Cells highlighted with colour shows data points which are not in the dependency distribution.

Table 6.1A: Milling Data Set Find Dependencies Table

	Lc		
V_f	**(-, 228.6)**	**[228.6, 635)**	**[635, +)**
(-,233)			
Predicted Time	28.15	142.7	43.87
Data points (cell)	100	83	42
Standard Error (cell)	79.14	361.5	293.6
Data points(sub-cell)	53	40	0
Standard Error (sub-cell)	62.41	296.4	--
[233,873.2)			
Predicted Time	0.4006	1.726	3.691
Data points (cell)	178	139	47
Standard Error (cell)	1.229	4.042	5.954
Data points(sub-cell)	176	111	47
Standard Error (sub-cell)	1.123	2.847	5.954
[873.2, +)			
Predicted Time	0.2027	0.3797	1.095
Data points (cell)	30	31	41
Standard Error (cell)	0.6094	1.294	2.103
Data points(sub-cell)	30	31	33
Standard Error (sub-cell)	0.6094	1.294	1.718

Figure 6.1A is a thermal chart which analyse the DS_2 in detail. In this figure, X-axis represent feed rate, Y-axis represent the milling process times and on Z-axis which is thick colored line represent machining length. This chart reveals that DS_2 has values with lesser feed rate and lesser cycle time in spite of having higher machining length. This is one of the exceptional features of DS_2. Similarly Figure 6.1B is a thermal chart for DS_2, in which tool diameter is compared against cycle time and feed rate. Figure A.2 reveals that majority of the data points in DS_2 has lesser tool diameter and lesser feed rate.

Figure 6.1A: Comparison of Process Variables for Milling DS_2 (T_m, V_f, and L_c)

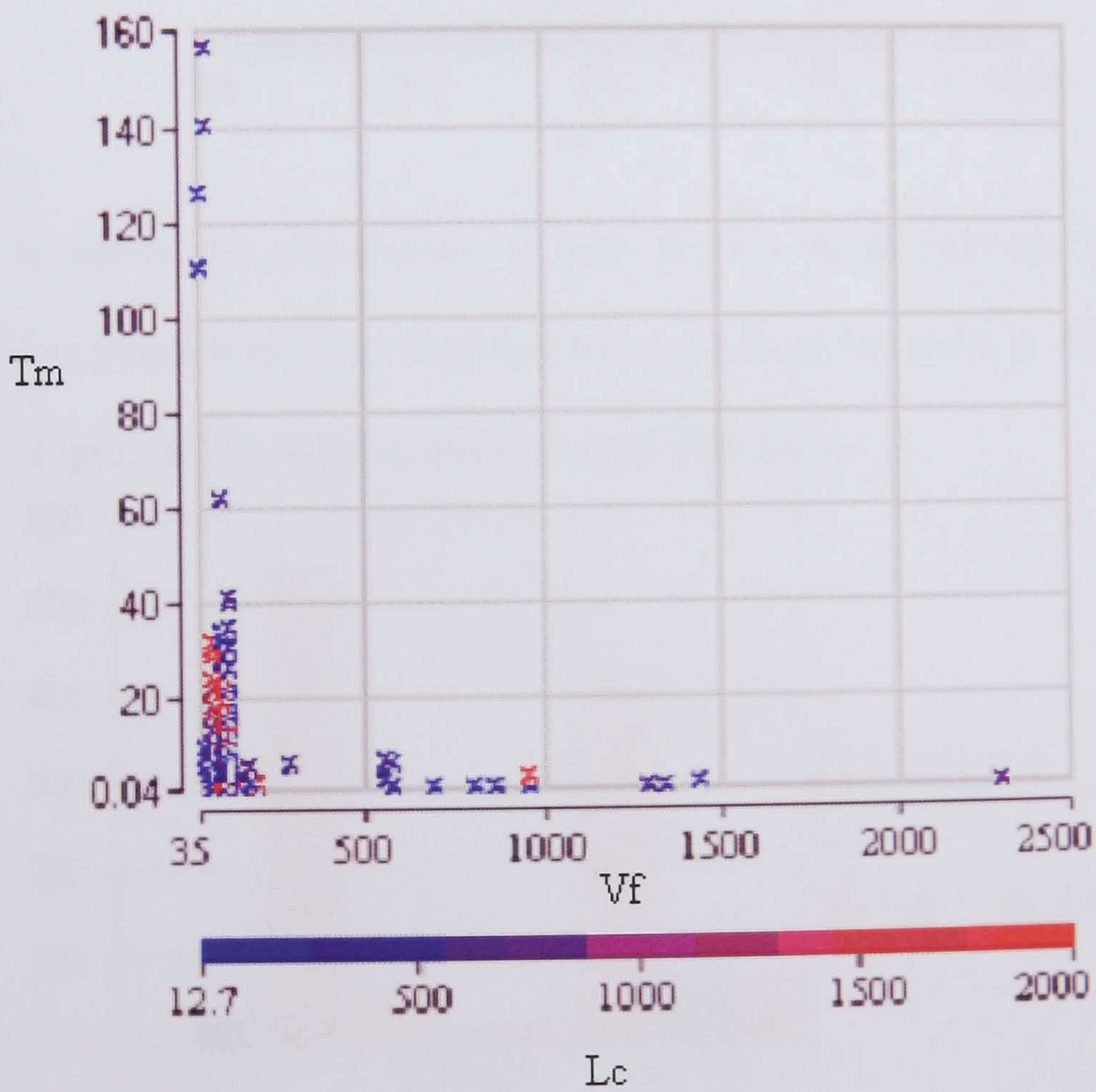


Figure 6.1B: Comparison of Process Variables for Milling DS_2 (T_m, V_f, and D)

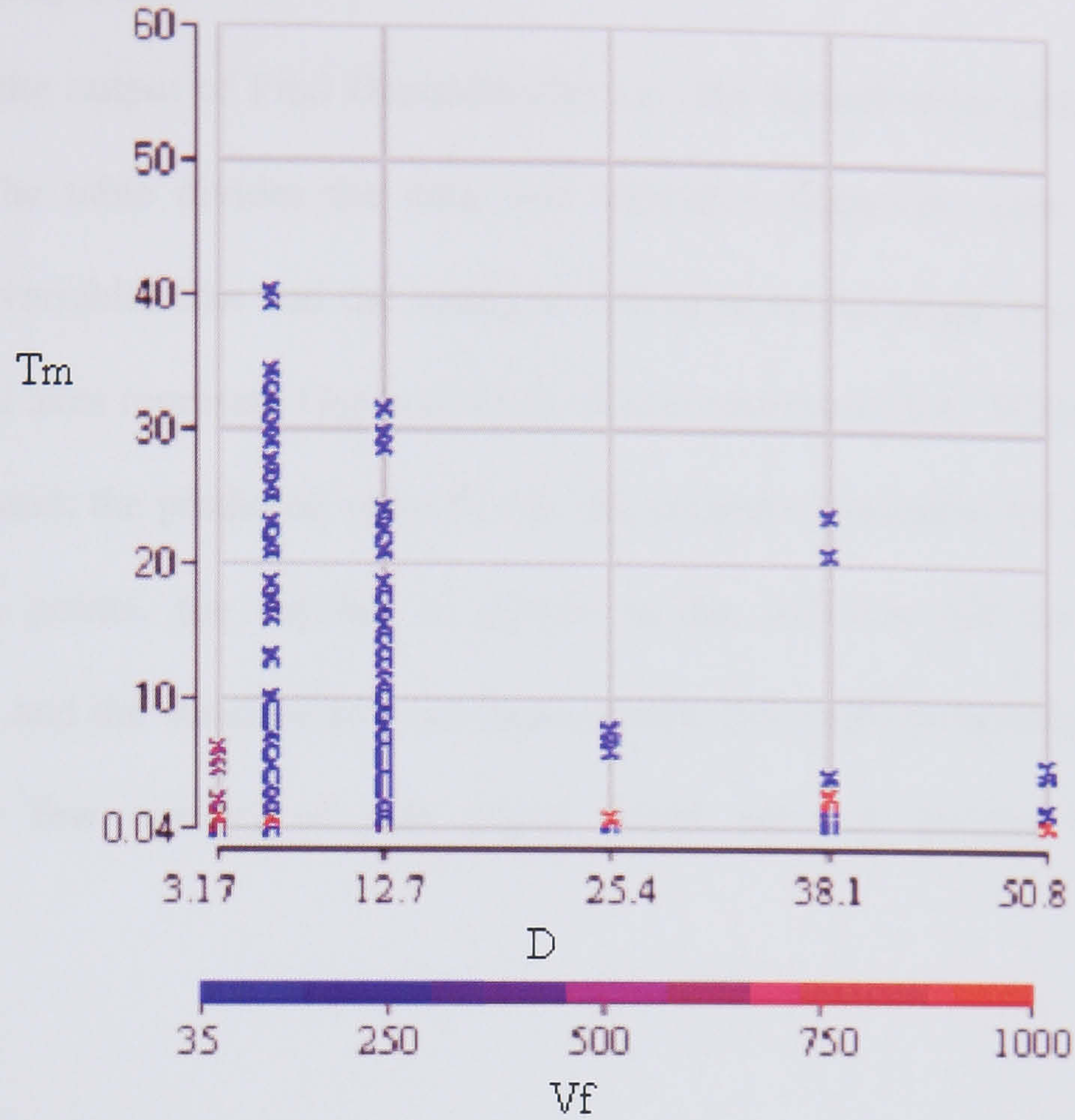
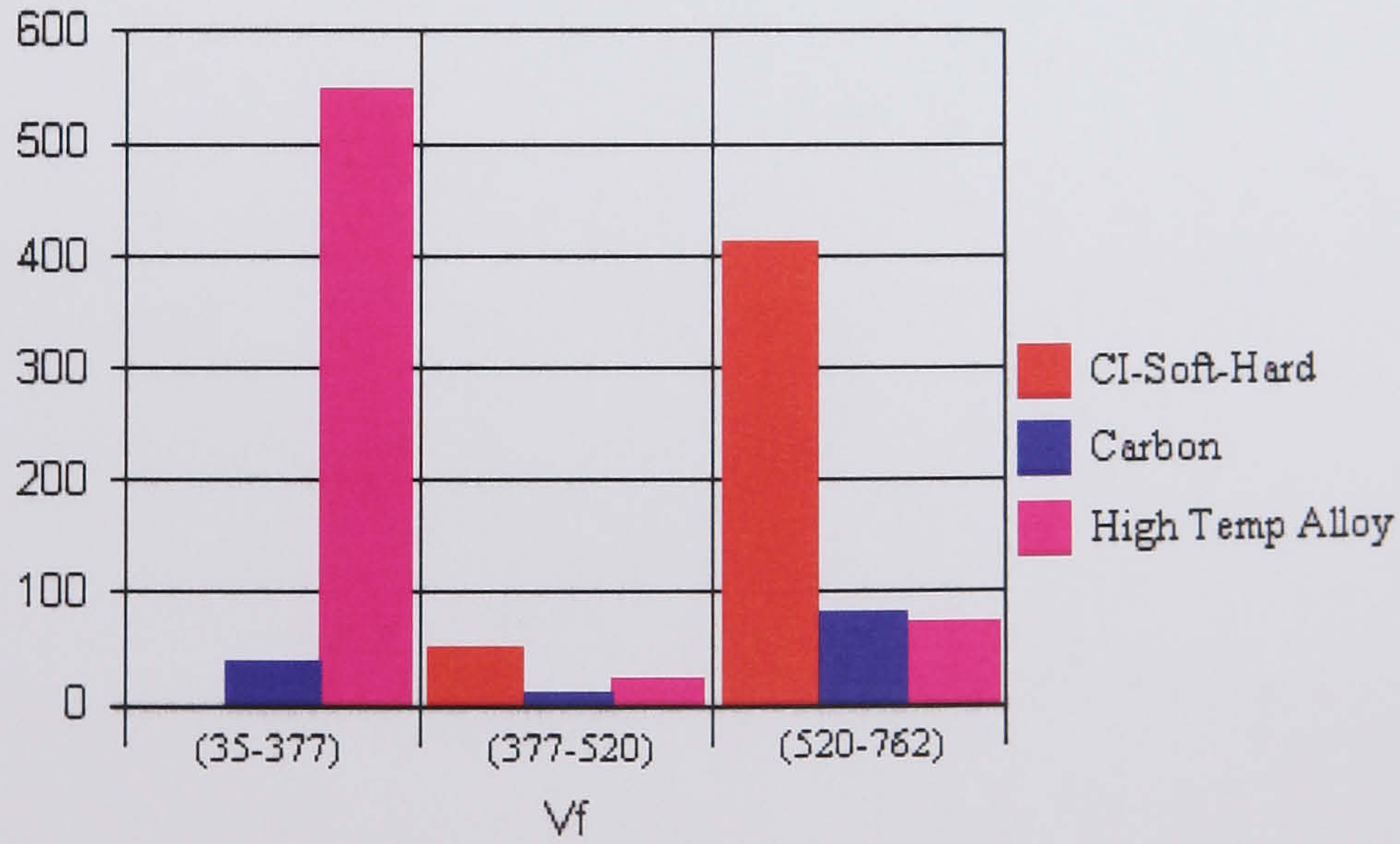


Figure 6.1C is shows the distribution of data points within different materials. This indicates that data points with lesser feed rate belong to High Temperature Alloys.

Figure 6.1C: Milling Data set and Materials



Appendix 6.2A

Automated Spray Paint

Table 6.2A is the output of Find Dependencies i.e., the dependencies table for Automated paint spray. The table divides the data into segments along two axes, these two axes represents the variables that had the strongest influence on the target dependent variables. Here horizontal axes represent Gsp and Vertical axes represent PthL. Within each cell, five numbers are listed; the predicted value (Ctm), the number of points in the cell, the standard error of those points, the number of points in the cell that fall into the dependent subpopulation, and the standard error of those points. It is evident form the table that there only are only few number of data points which are not obeying this dependency distribution.

Table 6.2A: Automated Paint Spray Find Dependencies Table

	Gs				
P_{tl}	(-, 400)	[400, 500)	[500, 600)	[600, 700)	[700, +)
(-, 1250)					
Predicted Time	2.407	**1.794**	1.621	1.408	1.254
Data points (cell)	38	**42**	60	45	51
Standard Error (cell)	20.02	**14.39**	13.98	11.03	10.27
Data points(sub-cell)	38	**42**	58	45	50
Standard Error (sub-cell)	20.02	**14.39**	12.43	11.03	9.803
[1250, 2300)					
Predicted Time	5.945	4.578	3.974	3.348	3.1
Data points (cell)	33	38	59	46	63
Standard Error (cell)	28.65	17.51	17.44	13.3	12.62
Data points(sub-cell)	33	38	55	42	63
Standard Error (sub-cell)	28.65	17.51	12.91	10.48	12.62
[2300, 3500)					
Predicted Time	9.156	7.731	6.602	5.714	5.424
Data points (cell)	34	40	53	46	68
Standard Error (cell)	44.97	23.42	21.68	18.75	19.03
Data points(sub-cell)	34	38	49	43	68
Standard Error (sub-cell)	44.97	21.45	15.2	13.18	19.03
[3500, 5000)					
Predicted Time	14.53	11.68	10.08	8.014	7.032
Data points (cell)	42	46	68	37	46
Standard Error (cell)	63.3	36.02	30.91	16.39	18.28
Data points(sub-cell)	39	41	68	37	46
Standard Error (sub-cell)	57.5	30.69	30.91	16.39	18.28
[5000, +)					
Predicted Time	21.14	16.34	15.71	12.27	10.62
Data points (cell)	51	32	58	24	69
Standard Error (cell)	98.54	67	61.36	41.68	32.55
Data points(sub-cell)	44	32	52	24	69
Standard Error (sub-cell)	81.86	67	58.83	41.68	32.55

Figure 6.2A and Figure 6.2B are thermal charts which analyse the DS_1 and DS_2 in detail. In this figure, X-axis represent paint path length, Y-axis represent the paint cycle times and Z-axis which is thick colored line represents paint gun speed. Figure 6.2A reveals that DS_1 has lower cycle time for the data points with higher gun speed (red colour) where lesser speed

has higher cycle time. Where as in Figure 6.2B shows that data points are behaving opposite to that of the points in DS_1 and is one of the reasons for DS_2 being different form DS_1.

Figure 6.2A: Analysis of Automated Paint Spray DS_1

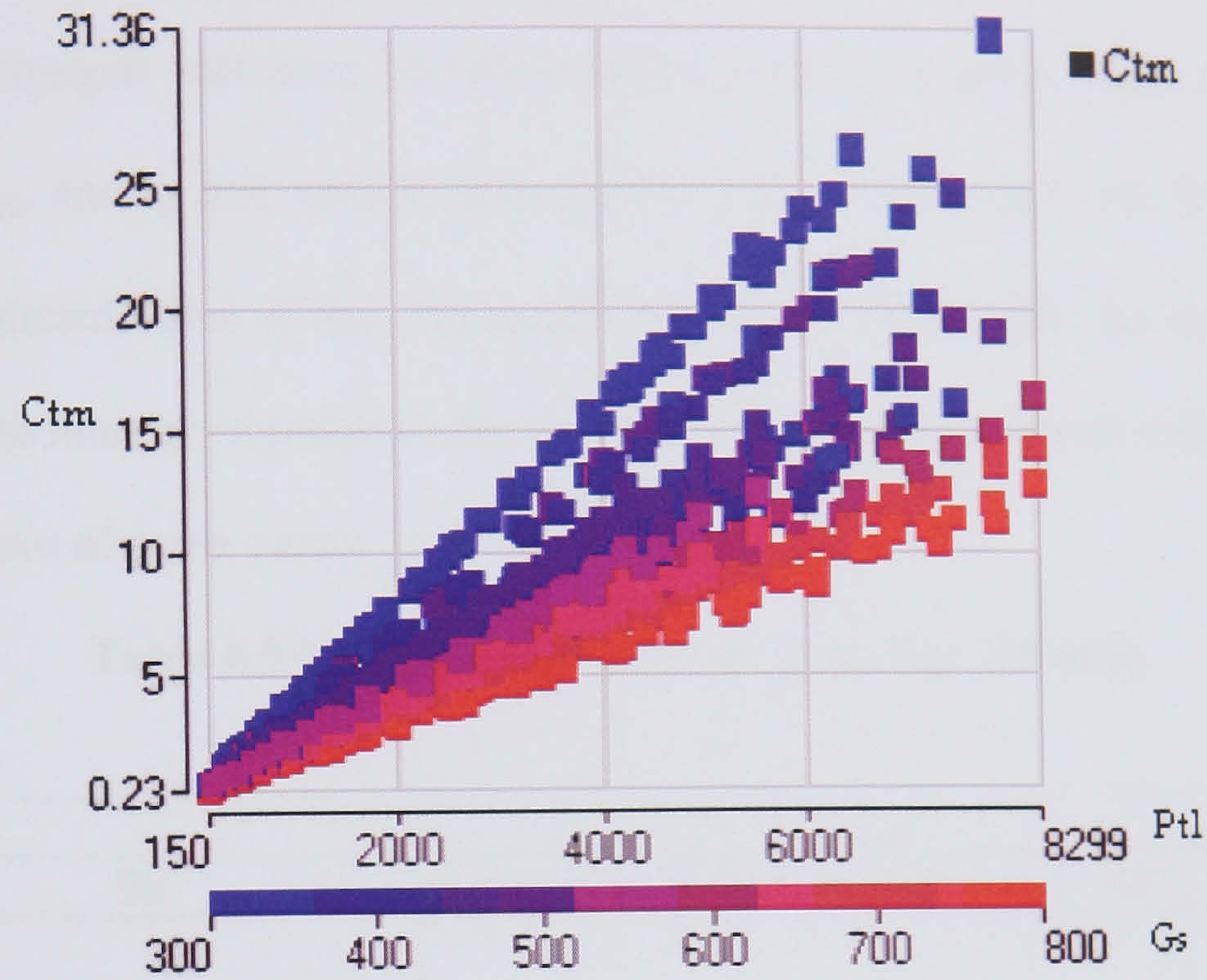
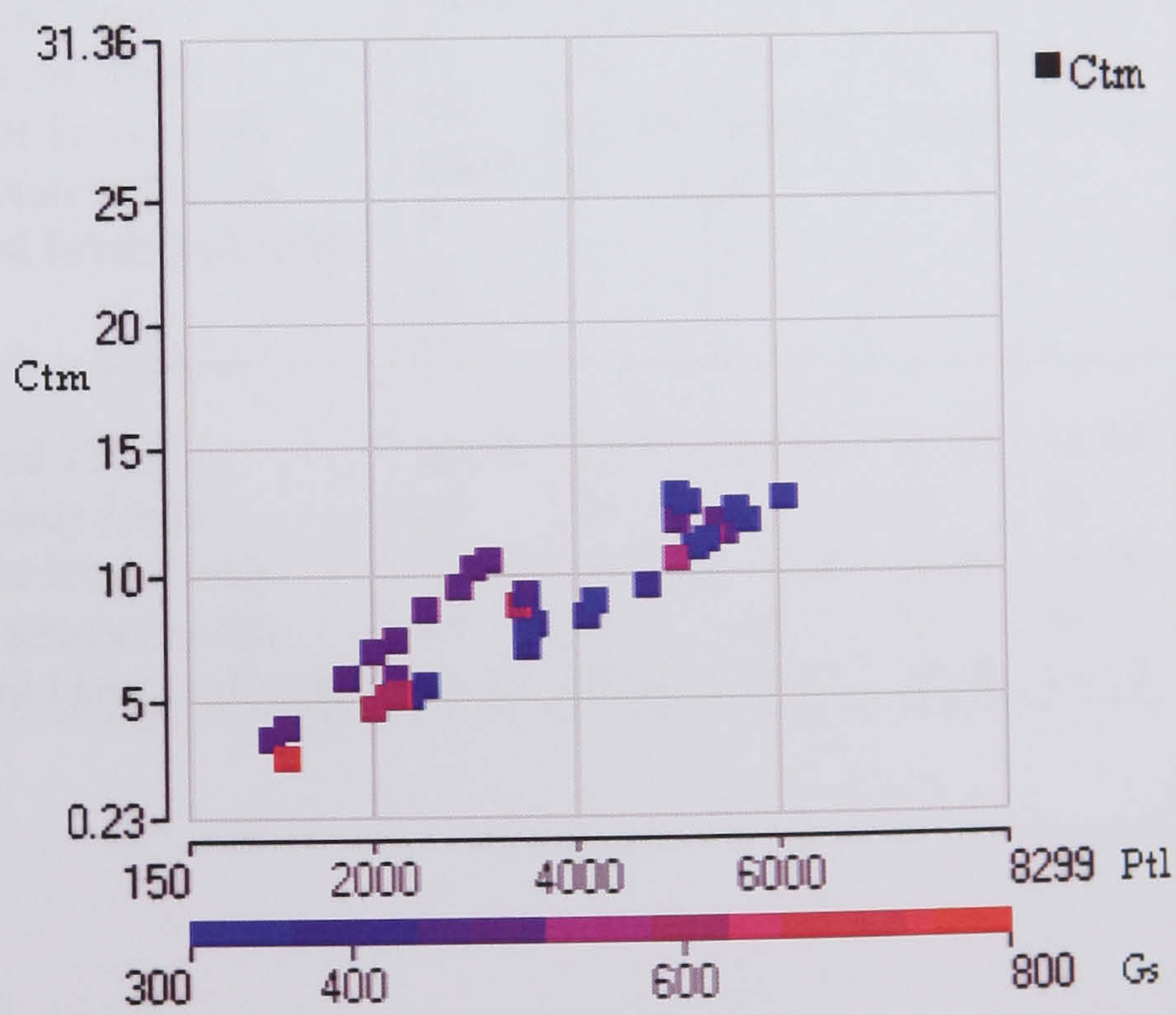


Figure 6.2B: Analysis of Automated Paint Spray DS_2



Appendix 6.3A

Turning

Table 6.3A is the output of Find Dependencies i.e., the dependencies table for Turning. The table divides the data into segments along two axes, these two axes represents the variables that had the strongest influence on the target dependent variables. Here horizontal axes represent n_t, t_{pt}, and t_{ln} and vertical axes represent B_s. Within each cell, five numbers are listed; the predicted value (Ctm), the number of points in the cell, the standard error of those points, the number of points in the cell that fall into the dependent subpopulation, and the standard error of those points.

Table 6.3A : Find Dependencies Table for Turning

	n_t		t_{pt}		t_{ln}	
Bs	**(-, 4)**	**[4, +)**	**(-, 6)**	**[6, +)**	**(-, 45)**	**[45, +)**
(-, 54) Predicted Time Data points (cell) Standard Error (cell) Data points(sub-cell) Standard Error (sub-cell))	115. 9 40 83.4 5 0 --	141.1 59 87.15 0 --	115.9 40 83.45 0 --	134.6 48 73.65 0 --	115.9 40 83.45 0 --	139.7 44 118.1 0 --
[54, +) Predicted Time Data points (cell) Standard Error (cell) Data points(sub-cell) Standard Error (sub-cell)	64.89 49 18.58 44 14.12	75.8 35 21.52 34 20.07	64.89 49 18.58 44 14.12	87.08 53 24.89 53 24.89	64.89 49 18.58 44 14.12	81.68 44 18.89 44 18.89

Figure 6.3A and Figure 6.3B are thermal charts that analyze the Turning DS_1 and DS_2. In this figures, X-axis represent basic setup times, Y-axis represent the turning cycle times and Z-axis which is thick colored line represents batch size.

Figure 6.3A :Turning Data Set (DS1)

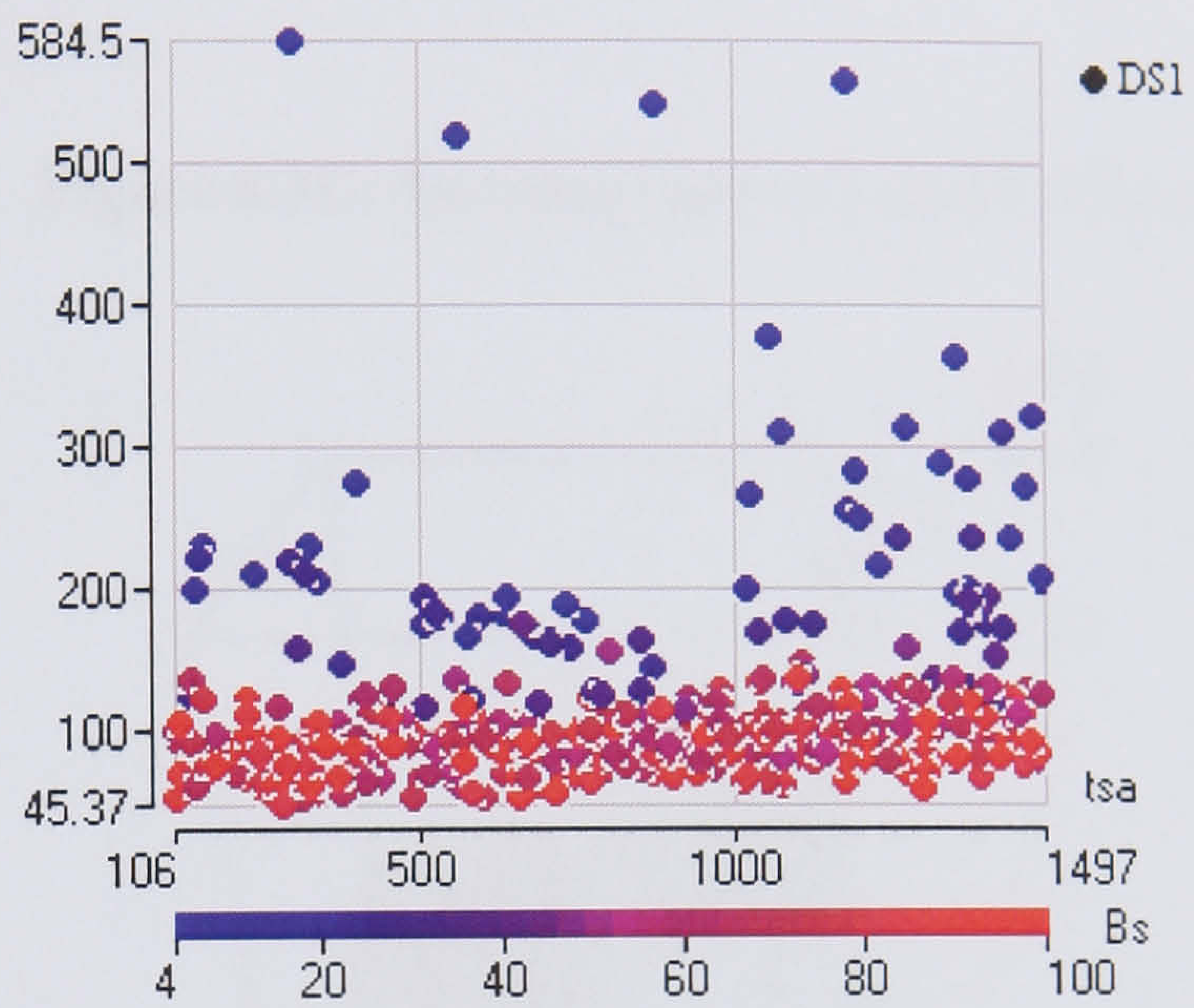


Figure 6.3B: Turning Data Set (DS2)

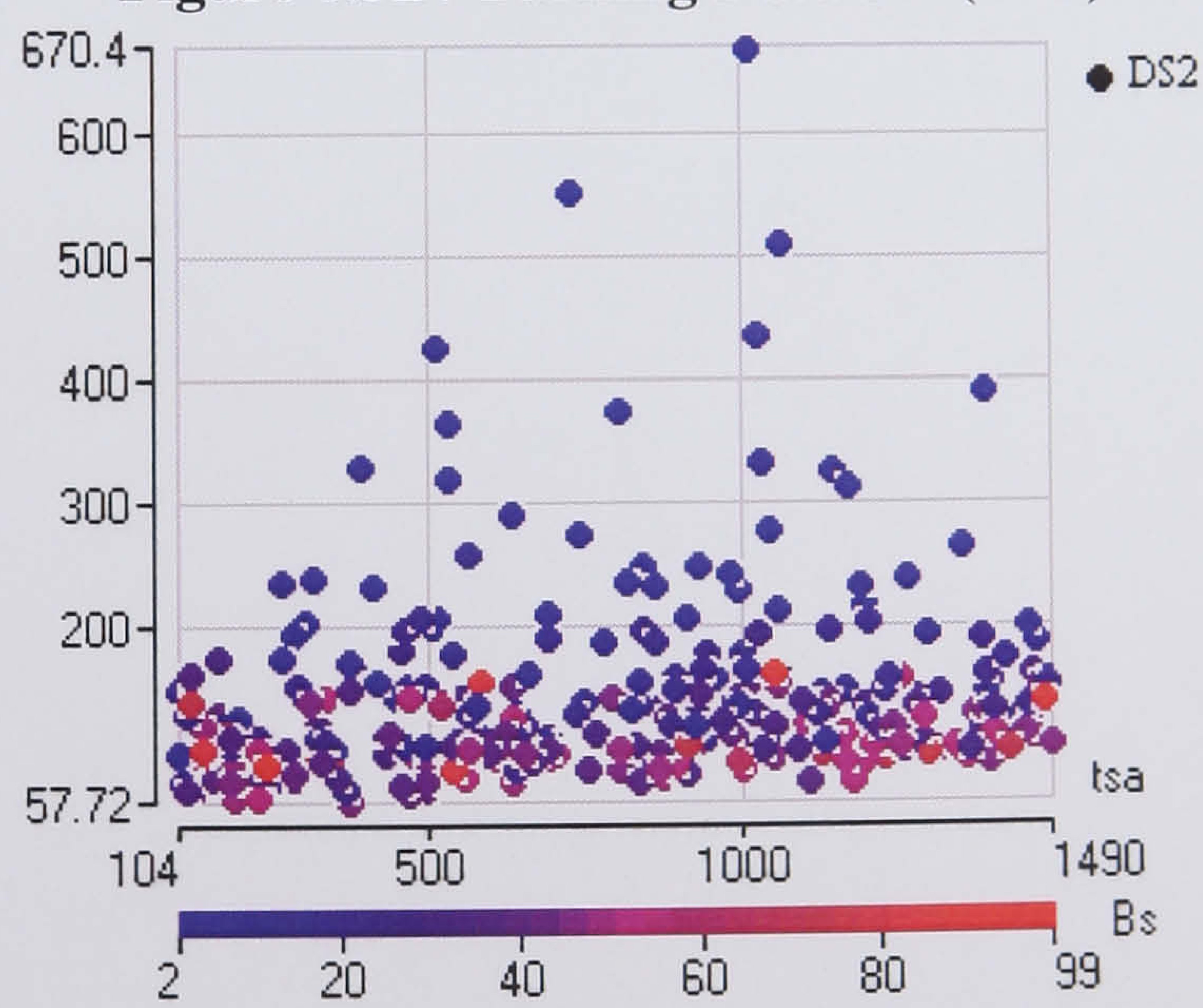
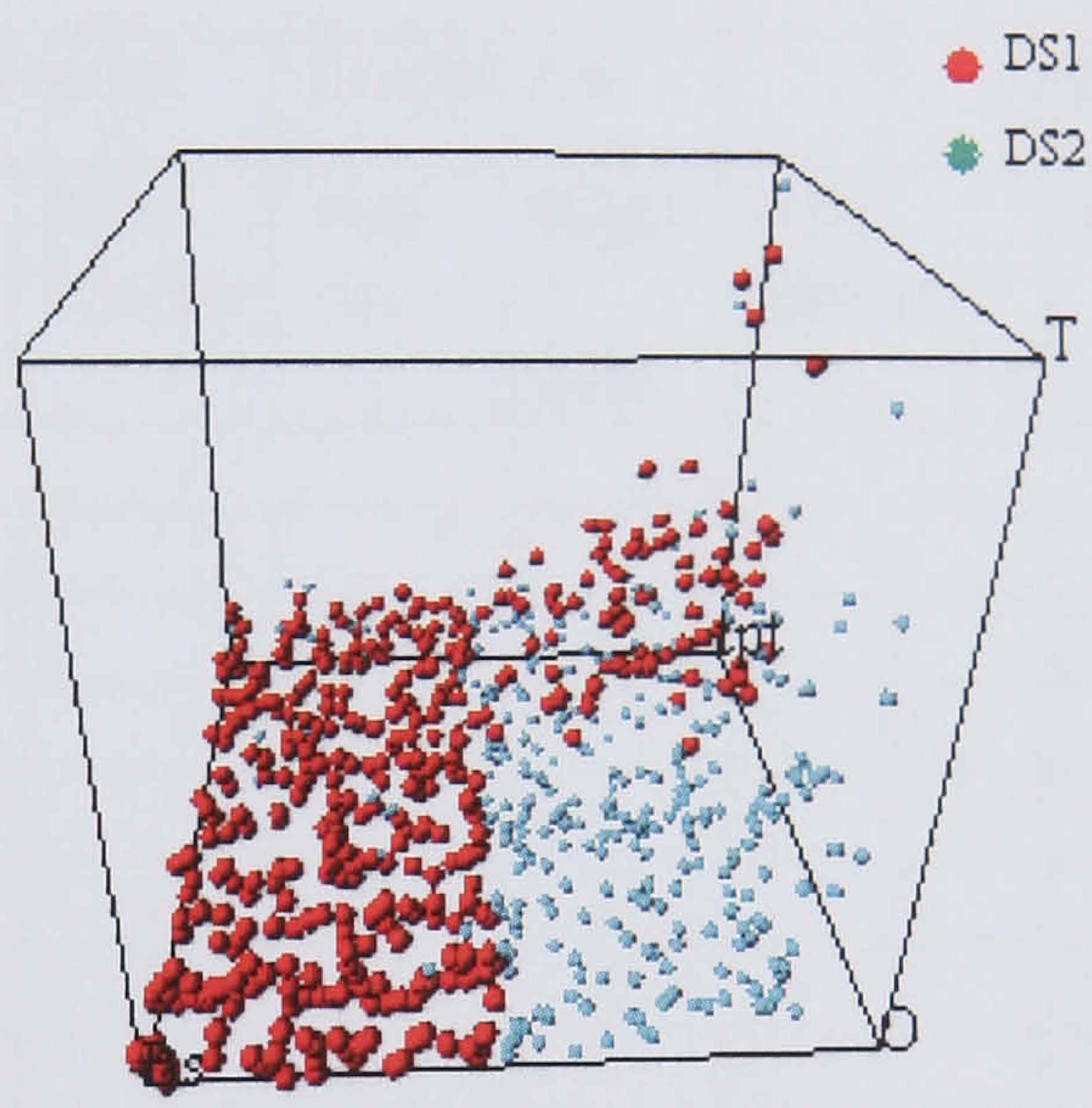


Figure 6.3C represents whole turning data points in a 3D charts. The x-aixs of this chart represents batch size, y-axis represents cycle time and z-axis represents too positioning time per operation.

Figure 6.3C: Turning Data Set on 3D Chart



Appendix 6.4A

Analysis of Predictive Models

Table 6.4A: Analysis of Vertical End Milling Models

	Actual					**Predicted**				
		Relationships		Coefficient			Relationships		Coefficient	
Variable Type	Cost Drivers	Linear	Non-Linear	+ve/ -ve	High / Low	Cost Drivers	Linear	NL	+ve/ -ve	High/ Low
						SLR				
D						√		√	-	**H**
$\mathbf{D_c}$						√		√	-	**H**
V_c						√		√	-	L
n						√		√	-	L
$\mathbf{V_f}$	√		√	**+**	**L**	√	√		**+**	**L**
$\mathbf{F_t}$	√		√	**+**	**L**	√		√	-	**H**
Nt										
$\mathbf{L_c}$	√	√		**+**	**L**					
						FL				
D										
$\mathbf{D_c}$										
V_c										
n										
$\mathbf{V_f}$	√		√	**+**	**L**	√		√	+	**L**
$\mathbf{F_t}$	√		√	**+**	**L**	√		√	+	**L**
Nt										
$\mathbf{L_c}$	√	√		**+**	**L**	√	√		+	**L**
						PnP				
D										
$\mathbf{D_c}$										
V_c						√		√	-	H
N										
$\mathbf{V_f}$	√		√	**+**	**L**	√		√	**-**	**H**

F_t	√		√	+	L					
Nt										
L_c	√	√		+	L	√	√		+	H

Table 6.4B: Analysis of Automated Spray Paint Models

	Actual					Predicted				
	Cost Drivers	Relationships		Coefficient		Cost Drivers	Relationships		Coefficient	
Variable Type		Linear	Non-Linear	+ve/ -ve	High/ Low		Linear	NL	+ve/ -ve	High/ Low
						SLR				
D_{wp}						√		√	-	H
G_s	√		√	-	H	√		√	-	H
G_{rg}						√		√	-	L
P_{tl}	√	√		+	H	√	√		+	L
P_{ns}	√	√		+	H	√	√		+	L
P_{nw}						√		√	-	H
P_{fr}	√		√	-	H	√		√	-	
						FL				
D_{wp}										
G_s	√		√	-	H	√		√	-	H
G_{rg}										
P_{tl}	√	√		+	H	√	√		+	L
P_{ns}	√	√		+	H	√	√		+	H
P_{nw}										
P_{fr}	√		√	-	H	√		√	+	H
						PnP				
D_{wp}										
G_s	√		√	-	H					
G_{rg}										
P_{tl}	√	√		+	H	√	√		+	H
P_{ns}	√	√		+	H	√	√		+	H
P_{nw}										
P_{fr}	√		√	-	H	√		√	-	H

Table 6.4C: Analysis of Turning Models

	Actual					Predicted				
		Relationships		Coefficient			Relationships		Coefficient	
Variable Type	Cost Drivers	Linear	Non-Linear	+ve/ -ve	High / Low	Cost Drivers	Linear	NL	+ve/ -ve	High/ Low
						SLR				
r_v	√	√	√	+	**H**					
ps	√	√	√	+	**H**	√	√		+	**H**
W	√		√	+	**H**					
d_m	√	√	√	+	**L**					
R_{sg}	√		√	+	**L**					
r_i	√		√	+/-	**H**					
l_r	√		√	+	**H**					
r_e	√		√	+/-	**L**					
n	√	√	√	+	**L**					
t_{sa}	√	√		+	**L**	√	√		+	L
n_t	√	√		+	**L**	√	√		+	H
t_{sb}	√	√		+	**L**	√	√		+	L
B_s	√		√	+	**L**	√		√	-	L
t_{ln}	√	√		+	**L**	√	√		+	L
n_o	√	√		+	**L**	√	√		+	H
t_{pt}	√	√		+	**L**	√	√		+	H
						FL				
r_v	√	√	√	+	**H**					
ps	√	√	√	+	**H**					
W	√		√	+	**H**					
d_m	√	√	√	+	**L**					
R_{sg}	√		√	+	**L**					
r_i	√		√	+/-	**H**					
l_r	√		√	+	**H**					
r_e	√		√	+/-	**L**					

n	√	√	√	+	L					
t_{sa}	√	√		+	L	√	√		+	L
n_t	√	√		+	L	√	√		+-	L
t_{sb}	√	√		+	L	√	√		+	L
B_s	√		√	+	L	√		√	+	H
t_{ln}	√	√		+	L					
n_o	√	√		+	L	√	√		+	H
t_{pt}	√	√		+	L					
								PnP		
r_v	√	√	√	+	H					
ps	√	√	√	+	H					
W	√		√	+	H					
d_m	√	√	√	+	L					
R_{sg}	√		√	+	L					
r_i	√		√	+/-	H					
l_r	√		√	+	H					
r_e	√		√	+/-	L					
n	√	√	√	+	L					
t_{sa}	√	√		+	L					
n_t	√	√		+	L					
t_{sb}	√	√		+	L					
B_s	√		√	+	L	√	√	√	+/-	
t_{ln}	√	√		+	L	√	√	√	+/-	
n_o	√	√		+	L					
t_{pt}	√	√		+	L					

Table 6.4D: Analysis of Data Sets

	% Distribution of data points					
Variable Type	Low	Med	High	Mean	Dispersion	Range
Vertical End Milling						
D	36	52	12	1.01	0.5886	1.87
D_c	42	20	38	1.279	1.004	2.92
V_c	62	23	15	325.8	266.1	940
n	88	10	2	2109	2753	30445
V_f	76	23	1	34.36	36.6	365.3
Lc	71	22	7	20.6	19.12	79
F_t	72	20	7	0.00959	0.01027	0.039
Automated Paint Spray						
D_{wp}	34	33	34	149.8	52.31	150
G_s	33	33	34	499.7	143.5	550
G_{rg}	66	25	9	540.3	153.5	500
P_{tl}	35	45	19	376.2	67.16	250
P_{ns}	87	13	1	3118	1994	8149
P_{nw}	84	15	2	7.407	5.412	268.5
P_{fr}	16	50	34	49.12	36.69	997.9
Turning						
r_v	38	21	41	0.6055	0.1351	0.4
ps	28	34	38	1.001	0.32	1
W	28	34	37	36.7	14.38	50
d_m	34	30	36	0.2025	0.08386	0.2

R_{sg}	33	35	32	123.6	41.98	150
r_i	63	33	4	0.1687	0.07409	0.25
l_r	18	61	21	2.527	1.434	4.9
r_e	29	31	40	0.5373	0.2583	0.8
n	26	48	26	0.3491	0.1133	0.3
t_{sa}	25	22	53	840.8	407.8	1391
n_t	31	20	49	3.477	1.701	5
t_{sb}	29	38	33	316.3	157.7	550
B_s	12	19	69	66.61	24.68	96
t_{ln}	34	31	36	45.61	9.185	30
n_o	31	35	34	3.602	1.615	5
t_{pt}	25	29	47	5.895	2.923	9